Class Debrief: A pilot study exploring students’ use of a video-based collaborative indexing and review system for lecture-based instruction
Paula Wellings (paulaw@stanford.edu)
June 5, 2004
Abstract

This pilot study investigates how students learning activities in large lecture-based instructional environments might be influenced by the use of a video-technology system that enable students to mark relevant parts of a lecture and then review the lecture through the use of their marks and the marks identified by other students. Of particular interest is learning about students identification of significant moments in a lecture, how students interact with moments in a video lecture identified by other people, and how students marking and review strategies relate to test taking performance. While this initial study is confounded by a number of issues including technical difficulties, sample size and selection, and test measure construction, tentative findings suggest that certain participant activities within the system impact test outcomes. Specifically, it was found that participants who marked and reviewed sections of the lecture aligned to test questions had higher test scores than participants who did not mark these areas or review marks related to these areas. These findings are significant because they mirror findings associated with traditional note-taking activities.

Introduction

The affordances of the large lecture classroom often create an environment premised on an ephemeral interaction between the lecturer as performer and the students as audience. The primary activity of students during the lecture is the recording of the lecture through note-taking. 94 % of U.S. college students regard note-taking as pivotal to assimilating lecture content (Dunkel & Davy, 1989), and “most students attempt to record lecture content even when they are not explicitly instructed to do so”. (Williams & Eggert, 2002). Note-taking to support student learning is premised on the belief that it fulfills both a process function and a product function (Kiewra, 1985). The process function suggests that note-taking facilitates generative processing of new information. However, this premise is debatable, both from the perspective that the information-processing demands of lecture learning are too high to allow for generative processing, and from the perspective that the content of notes taken are not reflective of generative thought. “ Peper and Mayer (1978, 1986) found that note-takers outperformed nonnote-takers on a problem-solving test. However, there was no relationship between the content of students’ notes and their performance on the problem solving test (in Kiewra &
DuBois, 1991). This finding raises the question of whether writing notes or simply noting relevant content lead to test performance results.

The second premise of note-taking is the product function, that notes taken can be reviewed to further extend learning. Bromage and Mayer (1986) have demonstrated that repeated exposure to text or lecture information results in greater learning. Kiewra and Dubois, additionally, suggest a third aspect of note-taking, the review of other students notes without attendance of the lecture. In a study conducted with three experimental conditions: take notes during lecture but no review of notes, take notes during lecture and review notes, and review notes of another without initial lecture, Kiewra and Dubois found that the group that reviewed the notes without the lecture outperformed the note-taking but no review group on the synthesis test. The synthesis test required generative processing in order to form relations that had not been stated explicitly in the lecture.” They believe that this finding demonstrates that generative processing is not occurring during note-taking, but rather is taking place during note reviewing.

Research suggests that the note-taking process is related to lecture learning, but what about this process specifically results in learning is unclear. Given this situation, it is of interest to consider methods for noting significant parts in lectures that do not lead of cognitive overload and continue to support both the processes and products associated with student lecture learning.

This pilot study is an initial investigation into how a new technology system and associated processes might influence students learning activities in large lecture-based instructional environments. The technology system, Class Debrief, enables students in lecture-based classes to identify key areas of a lecture they wish to reflect upon post-lecture and discuss these areas with other students and faculty through the use of video records and online activities. The system is premised on the idea that the large lecture classroom can be video recorded and indexed by students for post-lecture review.

The specific focus of this pilot study is on the aspects of the system associated with students marking relevant parts of a lecture and reviewing those parts and parts identified by other students in preparation for a test. Of particular interest is learning about students identification of
significant moments in the lecture, how students interact with moments in the video lecture identified by other people, and how students marking and review strategies relate to test taking performance.

General Description of Experiment

Overview of the manipulation and the measures
The goal of this study is to investigate what effect identifying segments of perceived relevance in a video lecture and the subsequent review of these segments and segments made by others can have on learning via the lecture method. To achieve this goal, this study has been designed to look for possible relationships between the parts of a lecture participants mark for review, their reviewing behaviors including both the use of their own personal marks and marks given to represent the input of other students, and their performance on a test that is indexed to specific parts of the video lecture.

The study is organized around three activities: the viewing and index marking of a video-based lecture, the review of the video lecture through the use of personal index marks and the marks of others, and the completion of a test with questions indexed to moments identified within the lecture. Study participants are divided into two conditions for the review activity. One group sees index marks oriented towards one half of the questions on the test and the other group seeing index marks orients towards the other half of the questions on the test. The reason for the two conditions is to expose the effects of marking and review activities in contrast to the questions and lecture content activating potential effects. By exposing each condition to different given marks, it is possible to measure the impact that marks, given to represent the input of other students, have on participant review activities and test performance.

It is hoped that findings from this pilot study will reveal if students marking of significant moments in a lecture aligns to test performance, if review activities impact test performance and if students utilize peer input as an effective learning resource. These findings have the potential to positively impact learning in large lecture classroom through the development of learning technologies that support these new activities.
Methods

Participants
Eight subjects participated in this pilot study. All participants were graduate students in Stanford’s School of Education, at the Master’s and PhD level and were selected on the basis of convenience and willingness to participate. As the lecture presented to study participants was related to online gaming communities, it is relevant to note that while all participants were familiar with the lecturer, Will Wright of The Sims Online, no participants were active in the Sims Online community or any other online gaming communities. Two subjects were removed from the data analysis due to excessive technical difficulties during the experiment. Two conditions were constructed to counter-balance exposure to review materials. One condition included two males and one female; the second condition included three females. Participants ranged in age from their 20s to 40s.

Materials

SCPD Video
All participants in the study watched an 11 minute excerpt from a talk given by Will Wright to a guest lecture class, CS 547: Human-Computer Interaction Seminar (Seminar on People, Computers, and Design). This class is video taped and produced for online streaming by the Stanford Center for Professional Development (SCPD). Presentation slides are intermixed with shots of the lecturer speaking. In this particular video, Wright utilizes many diagrams that in final production are shown in the video with Wright’s voice over.

The 11 minute excerpt focuses on how the online community associated with the Sims Online functions and details the roles that community members play in relationship to the game and to each other. This video segment was selected as it presented a self-contained unit of ideas with both factual and conceptual attributes within a reasonable length of time, accessible to an audience of computer-literate study participants.

Class Debrief marking interface
All participants watched the video excerpt in a computer interface. The interface played the video from beginning to end and provided participants with a single button to press during
viewing. Participants were instructed to press the button when they wished to mark a point in the video that they would like to return to for review. Due to technical difficulties, this button did not always mark places in the video for review as expected, but participants were not aware of this malfunction while using the Class Debrief marking interface. *(see Appendix B)*

*Class Debrief review interface*

Participants in the two conditions viewed two versions of the Class Debrief review interface. Both versions of the interface presented the participant with the marks they had made while watching the video on a video timeline in green. In addition, both versions presented participants with marks in red represented as being made by nine undergraduate computer science students who attend CS 547. In reality, these marks were manufactured to align with parts of the video associated with questions on the test. One condition saw marks associated with seven of the 14 questions, and the other condition saw marked associated with the other seven questions. The red marks were placed on the time line to suggest some individuality in placement, but with obvious visual aggregation. *(See the appendix to see the layout of the Class Debrief review interface in the two conditions.)*

Using the Class Debrief review interface, participants reviewed the video lecture by clicking on areas on the video timeline. Participants were able to move back and forth in the lecture and pause and play the video. Participants in both conditions used the Class Debrief review interface for four minutes.

*Test*

All participants spent up to 12 minutes taking a test on the content of the lecture. The 14 questions of the test align with 14 identified segments within the 11 minute video. The test is composed of a number of question types including multiple-choice, fill in the diagram, draw a diagram, order given words, describe a procedure, and select characteristics. As each condition was presented with marks associated with 7 of the 14 segments, it is important that there is an equal distribution of question types associated with the seven segments from condition one and the seven segments from condition two. *(See appendix C.)*
Design
The design of the study focuses on three aspects of student activity: mark making, using the input of others to guide review, and more broadly, the impact of reviewing the video lecture through the use of number of types of index marks.

Mark making
This component of the study is designed to account for the total numbers of marks made by the participant and to determine the number of marks made that relate to the 14 questions on the test. The number of related marks is compared to the total number of marks made to account for the increased probability of marking a related place as more marks are made by the participant. The number of relevant marks made by the participant is also considered in comparison with the total number of relevant marks (14). It is expected that this design will reveal that participants who mark parts of the video that are related to test questions will perform better on the test.

<table>
<thead>
<tr>
<th>Marks</th>
<th>Total marks</th>
<th>Total number of marks made by participant during viewing of video lecture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marks</td>
<td>related</td>
<td>Total number of marks made by participant corresponding to questions +- six seconds</td>
</tr>
<tr>
<td>Marks</td>
<td>% related</td>
<td>Related marks divided by total number of marks made by participant</td>
</tr>
<tr>
<td>Test</td>
<td>% correct</td>
<td>Number correct divided by total number of questions</td>
</tr>
<tr>
<td>Marks</td>
<td>% marks identified</td>
<td>Related marks divided by total number of marks related to questions (14)</td>
</tr>
</tbody>
</table>

Using the input of others to guide review: given mark versus not given mark in review
This component of the study is designed to enable comparison making between performance on questions where the participant is given question relevant marks in the review and when the participant is not given question relevant marks. There are 14 questions associated with locations on the video timeline. Each condition is able to see one set of seven marks represented by marks made by the nine invented computer science students. This measure is dependant on an equal distribution of difficulty across all test questions, or at a minimum an equal distribution of difficulty of the questions between the two conditions.
The impact of reviewing the video lecture through the use of number of types of index marks: mark reviewing

This component of the study is designed to look at the test outcomes based on how participants review the video lecture in the four minutes review session. Looking specifically at marks aligned to test questions, marks within the review session are divided into four types:

- **given mark**: marks given under the supposition that these are parts of the video identified by computer science students
- **personal mark**: marks made by the participants during the video lecture viewing session
- **both**: personal marks that align with given marks, plus or minus six seconds
- **no mark**: points in the video lecture timeline correlated to test questions that are not marked by personal marks made or by given marks

These marks are then related to whether they are reviewed within the review session and whether or not the correlated test question is correct or incorrect.

It is expected that this design will reveal what types of marks and what review strategies are related to successful test performance.
Procedure
The study begins with a brief questionnaire to determine if the participant is familiar with the subject matter of the video lecture. All participants are then given an overview of the study, including the manufactured premise that the lecture was also shown to undergraduate computer science students who marked parts they found to be relevant. All participants watch the 11 minute video, marking points they find relevant. Participants in condition one are then shown a review screen that includes their marks on a video time line as well as marked areas that correspond to seven of the test questions. Participants in condition two are shown a review screen that includes their marks on a video timeline and the marks corresponding to the other seven test questions. Participants are able to review the video for four minutes. Participants are then given 12 minutes to complete a 14 question test. At the conclusion of the test, the study is complete. (See full study script, Appendix A)

Coding
Data collection methods for this study included the video capture of all on-screen interactions by participants and a 14 question test. Video data of the marking component of the study was reviewed to confirm the time marking of video segments by participants. A listing of participant actions in the review component of the study was created to facilitate the coordination of marks made by the participants, marks given to the participants, marks reviewed and not reviewed by the participants and their test performance. Participants’ marks were considered aligned with given marks if they fell within six seconds on either side of the mark. Through informal content analysis of the video, it was found that this range approximately contained a given unit of content aligned with a test question. Each of the test questions was scored correct or incorrect. Test scoring was not adjusted to account for variation in question difficulty or the probability a particular answer being correct or incorrect.
Results

Overview of results

Results of this study suggest that certain participant activities within the Class Debrief system impacted participant test scores. Generally, it was found that participants who marked and reviewed sections of the lecture aligned to test questions had higher test scores than participants who did not mark these areas or review marks related to these areas. However, it was not possible to conclude whether marks representing the perspectives of other students positively or negatively impacted participant test performance.

Details and graphs

Making marks

Figure one

Figure one shows the relationship between the number of test questions participants got correct and the percent of the marks that the participants made that were directly related to questions. For example, the study participant represented furthest to the right made a total of 10 marks; five of these marks were aligned to questions on the test. This graph suggests that participants with a lower percent of relevant marks to total marks made scored lower on the test, while participants with a higher percent of relevant marks scored higher on the test.
Figure two

Figure two shows the relationship between the number of test questions participants got correct and the percent of the related marks that the participants made of the total number of related marks. For example, the study participant represented furthest to the right marked five of the 14 marks aligned with test questions. This graph suggests that participants with a lower percent of relevant marks to total marks related to the test scored lower on the test, while participants with a higher percent of relevant marks scored higher on the test.
Figure 3 shows the distribution of test performance based on marks given and not given in the case of each condition. Participants in condition one answered more questions correctly that were associated with a given mark. Participants in this condition also answered more questions incorrectly when not given question associated marks. In contrast, participants in condition two answered more questions correctly that were not associated with a given mark and less questions correctly that were associated with a given mark. Overall, this graph does not show any conclusive evidence for a positive or negative test outcome based on marks given or not given.
Reviewing marks

![Distribution of test performance based on marking and reviewing](image)

**Figure 4**

Figure four shows the distribution of test performance for both groups based on marking and review behaviors. Represented on this graph is the number of questions correct and incorrect, whether or not marks were reviewed, and whether marks were given, personal, both given and personal, or there was no visible mark associated with a question. When looking at this graph it is important to note that for each condition seven marks associated with questions were given and seven marks were not. This becomes meaningful when looking that the bars that represent the percent of total questions answered related to no marks. Unless the participant made a personal mark in alignment with the not-apparent seven marks associated with questions, the responses to these questions would be identified as falling under the no mark category. It is also highly improbable that areas without visible markers will be reviewed.

Figure four shows that marking and review behaviors have a positive impact on test performance. Personal marks and personal marks aligned with given marks yielded a greater
percentage of correct answers than marks given. This can be seen more clearly in Figure 5 below.

![Personal marks compared with given marks](image)

**Figure 5**

In Figure 5, personal marks are combined with “both” marks—personal marks that are also aligned with given marks. In this view it is possible to see that marks made by the participant have a greater impact of question performance than marks “given” by the experiment representing the marks made by manufactured computer science students.

**Discussion**

**Issues with results**

Before discussing the potential meaning of these results, it is important to note that the results of this pilot study must be viewed with some suspicion. A number of factors in the construction and conduct of this study lead to the conclusion that the results cannot be considered wholly valid. Significant issues include technical problems with the marking interface, sample selection and size, and design of the test measure.
Technical problems with the marking interface

Participants were asked to click on a green button to mark parts of the video lecture they were interested in reviewing in preparation for a test. It was found halfway through running participant in the study that the green button had a malfunction. Button clicks were not recorded if the participant did not remove their mouse from the button between clicks. At this time, participants were instructed to remove the mouse from the button between clicks. The impact of this malfunction was that, when participants went to the review interface, not all of their marks were visible. It was possible to account for all button clicks through review of the video and to account for the number of marks each participant was missing in their review. Participants included in the study missed between 0 and 50% of their own marks in the review session. It was hoped that the study would reveal the impact of reviewing ones own marks, but technical difficulties cast this relationship into question. However, it is still possible to see the relationship between marks made and test performance, as all marks made have been accounted for through review of screen capture video.

Sample selection and size

The initial size of the study was four participants per condition. Due to excessive technical problems, two participants were removed from the study. Given the complexity and variability in experience, note-taking and listening strategies, three participants per condition cannot overcome potentially significant individual differences.

Design of the test measure

The test measure was designed to measure conceptual understanding as well as recall. While the test may have accomplished some of this goal, it was poorly designed from the perspective of balanced and distributed difficulty. As it turned out, the questions aligned to condition two’s given marks were more difficult from both the perspective of complexity and probability of selecting the correct answer. It was hoped that the study would measure the value of being given marks related to particular test questions, with the ability to counter-balance between condition one and condition two. The variability of test questions between the two groups makes this comparison of questionable validity.
Discussion of results

Making marks
The results most resilient to the issues stated above are those related to mark making. The mark making finding do not reflect the technical issues expressed in the review component of the study. These finding represent all eight participants overall performance on the test, although obviously the test measure is collected after the review which differentiates given marks between the two conditions. The Making marks component of the study suggest that students who successfully identify parts of the video that correlate to test questions will perform better on those questions than students do not identify these parts of the video.

Given mark versus not given mark in review
The Given mark versus not given mark in review component of the study likely suffers from the uneven design of the test measure and the small sample size. It is therefore difficult to conclude whether given marks influence test performance in a negative or positive way. The findings from the two conditions do suggest however test performance is influenced. It may be worthwhile to improve this component of the study in order to gain a better understanding of the value of students seeing places in the lecture identified by their classmates or by their instructor.

Mark reviewing
The Mark reviewing component of the study suggests that mark making and review have a positive impact on test performance. Based on this finding, it may be of interest to compare this mode of marking and review to traditional note-taking methods and to more rigorously explore the potentially generative aspect of mark making and review.

Outstanding issues
In this study it was found that students who marked and reviewed sections of the lecture aligned to test questions had higher test scores than students who did not mark these areas or review marks related to these areas. The question that is not answered by this study is why some participants identified areas aligned to test questions and others did not. This study also does not account for why these participants had higher test scores.
General Discussion

The focus of this pilot study was to learn more about students identification of significant moments in a lecture, how students interact with moments in the video lecture identified by other people, and how students marking and review strategies relate to test taking performance. By gaining a deeper understanding of these activities, it is hoped that student learning in large lecture learning environments can be broadened and improved.

While this initial study was confounded by a number of issues relating to technology, test design, and sample size, the finding do provide hints towards areas for further investigation. Specifically, the Mark marking component of the study reveals participant knowledge during the lecture and relates this knowledge to test performance. This finding mirrors C.C. Crawford’s pioneering research on the effects of note-taking in the 1920s: “Although Crawford obtained generally positive correlations between the number of ideas recorded in personal notes and the number included on a quiz, correlations between total points correct in both the notes and the quiz were consistently much stronger (in Williams & Eggbert, 2002).”

The Given mark versus not given mark in review component reveals that participants do interact with marks presented as belonging to other students. The Mark reviewing component reveals that the activities of mark making and mark reviewing impact test performance.

The study design issues identified above provide an initial guide for ways in which this study can be improved. To increase the validity of this study, technical issues in the software must be resolved, the sample must be increased and ideally be representative of student who attend large lectures, and the test measure must be revised to achieve a balance in question difficulty and correct answer probability.

In addition to these study design issues, other ways to improve and extend this study might include making the social aspect of the study more authentic with given marks representing real student mark making; and potentially separating the social aspects of review from the more individual aspects of review. To further explore the relationship between marking a place on the
video time line and traditional note-taking, it would be interesting to contrasting this video method of review with conventional note taking methods.

An improved study may have implications for theories and practices associated with the process and products of teaching and learning within the large-lecture learning environment. The lecture method has been the most common mode of instruction at the post-secondary level for the past 2000 years, and there are no signs that this will change in the near future. As such, it is a stable environment with much potential for research-based improvements and technological intervention.
References


Appendix of Materials

Appendix A: Study Protocol

1. Brief questionnaire

2. Three minutes introduction:

This is a study to determine the effectiveness of a new tool created for students to use in large lectures learning environments.

I will be showing you an 11 minute segment from a lecture series presented in by the computer science department. The segment is from a talk Will Wright, the founder of the Sims gave about the Sims Online. The segment you will be watching is specifically about online community.

Using the tool and interface I will introduce you to, you will be able to mark moments in the video of the lecture that you might like to watch again in preparation for a test which will follow.

Before beginning this study, I have shown this video segment to a number of students who attend CS 547, the lecture series this excerpt is from. I have also asked these students to identify the parts of the lecture they found to be most relevant for future reflection.

At the completion of the lecture except, the areas you identify will be available to you, as will be the areas identified by the CS students.

You will be able to review the video for 4 minutes.

Following your review of the video, I will give you a test that you have up to 12 minutes to respond to.

3. Introduce marking interface. Provide simple explanation.

4. View 11 minute video in marking interface.

5. Go to review screen. Provide simple navigation explanation.

6. Participant reviews video for four minutes.

7. Participant has up to 12 minutes to complete test.

End of study
Appendix B: Computer interfaces

Watch and mark video
Condition one: review video

Condition two: review video
Appendix C: Pre-study questionnaire:

Background Questions

Name:

Have you played the Sims Online?

Have you ever played an online video game? If yes, which ones?

Are you familiar with Will Wright, the founder of the Sims and the Sims Online?

Have you heard Will Wright speak before?

Have you read any interviews with him?

Do you participate in any online communities? If yes, which ones?
Appendix D: **Post-lecture test**  
Conditions are included here for reference, but were not included in test distributed to participants.

*You have 12 minutes to complete this test. Feel free to jump between questions.*

**CONDITION 1 | Question 1:**  
Circle key characteristics of the Sims game play (can circle more than one):

- competition  
- collecting  
- drama  
- creativity  
- storytelling  
- learning  
- imagination  
- skill

**CONDITION 1 | Question 2:**  
Put the following roles in order based on their hierarchical role within the community:

webmaster    casual player    collector    creator

[a] [b] [c] [d]

**CONDITION 2 | Question 3:** Circle the relationship between game play and community involvement as people progress from newbie to veteran:

- a. game play increases, community involvement decreases  
- b. game play increases, community involvement increases  
- c. game play decreases, community involvement increases  
- d. game play decreases, community involvement decreases
CONDITION 2 | Question 4:

What does this diagram depict?
   a. the number of people who change status within the community
   b. the number of websites the players browse before successfully adding items to their collections
   c. the number of players who win the Sims at each level of difficulty
   d. the number of people take on greater community responsibility and the who drop out

CONDITION 2 | Question 5: People engage in the Sim’s Online community to gain recognition. Complete these statements with the most likely group:

Story creators seek recognition from: _______________
Content artists seek recognition from: _______________
Tool Builders seek recognition from: _______________

webmasters      collectors      tool builders      browsers      content artists      story creators
CONDITION 2 | Question 6:
Put the following words in to three groups based on the content of Will Wright’s lecture:

learning        drama       story       hobby       sport       discovery       skill       creativity
competition

Group 1:          Group 2:          Group 3:

CONDITION 1 | Question 7: What is the impact of a thematically-specific game on community participation?
               a. people are better able to find their niche and gain recognition
               b. people are better able to find their niche but not gain recognition
               c. people are less able to find their niche and gain recognition
               d. people are less able to find their niche but able to gain recognition

CONDITION 1 | Question 8: Draw the shape of the community for the video game Half-Life.
**CONDITION 1 | Question 9:**
Put the terms on the dotted lines of the following diagram:

![Diagram with terms](image)

**CONDITION 1 | Question 10:** How does the Sims Online deal with competition from community websites?

- a. they keep upcoming game releases a secret to prevent community sites from publishing new materials ahead of time
- b. they tolerate it as long as the websites are not-for-profit
- c. they avoid competing with community sites that distribute game content
- d. they block web sites that distribute player-made game content

**CONDITION 1 | Question 11:**
According to Will Wright, which of the following is not open to creative interpretation?

- a. hobbies
- b. sports
- c. stories
- d. communities
CONDITION 2 | Question 12: Fill in this diagram:

CONDITION 2 | Question 13:
Provide a short description of how player stories are created?

CONDITION 2 | Question 14:
First person shooter video games are most like:
   a. stories
   b. hobbies
   c. sports