The Effect of Announced/Unannounced Examinations on Student Retention

John C. Reis
Aerospace Engineering
Embry-Riddle Aeronautical University
Prescott, AZ 86301

ABSTRACT

A hypothesis that students given unannounced quizzes in engineering classes, instead of announced midterms, would have a improved long-term retention of the course material was tested using two sections of an introductory fluid mechanics class. The hypothesis was based on the assumption that a series of unannounced quizzes minimizes "cramming" and results in more material being placed in long-term memory. Long-term retention was found not to be statistically different for the two sections and the hypothesis was not proven. When learning and testing occur at high levels in Bloom's taxonomy, substantial material is automatically placed in long-term memory and retention is independent of testing method, regardless of how much "cramming" occurs.

INTRODUCTION

Most engineering science classes evaluate student understanding of the material through one or more examinations during the semester and a comprehensive final examination. Many students prepare for announced examinations through an intensive study immediately before the examination, e.g., cramming. Although this method has proven to be a successful method for preparing for individual examinations, its value for long-term retention is questionable. Cramming normally places the learned material in short-term memory, where it is available for the imminent examination, but is forgotten shortly thereafter.

It was hypothesized that students would retain an understanding of the material for a longer period of time if their study habits were altered so that they studied the same material multiple times over the course of the semester instead of just prior to announced examinations. This behavior would place more of the material in long-term memory, improving retention for use in subsequent classes and after graduation.

A number of methods exist to encourage students to study the same material multiple times during the semester. Multiple examinations covering the same material would, at a minimum, force students to "cram" the same material a number of times. Multiple examinations, however, have the limitation that they replace classroom learning time with evaluation time. This loss of learning time limits the number of in-class examinations that may be given. Another method for encouraging students to study the same material multiple times is to give unannounced examinations. Not knowing which day the examination will be given encourages students to be more prepared at all times. Thus, they will likely review the material more times.
In one study, students in three sections of an elementary economics class were evaluated with three different testing procedures.1 One section was told at the start of the semester that they would be given tests periodically during the semester, but were not told precisely when. The second section was given the additional information that they would be given a test "soon" and the test was then administered from between 2 and 7 days after that announcement. The third section was told in advance the exact date of the tests. The section with no advanced warning performed significantly better than the other sections. No significant difference in performance was observed when students were told that the test would occur "soon" and when they knew exactly the date of the tests. Further, class attendance was higher for the section having no advanced warning.

Bahrick2, Conway et al.3, and Bahrick and Hall4 reported mixed results connecting course grade and long-term retention. Conway et al.5 and Bahrick6 reviewed these differences and suggested that course grades may not reflect long-term retention because grading methods can reflect different types of learning through the course. They reported that performance on course work was more strongly correlated with long-term retention than was final examination scores. It is noted however, that the types of courses evaluated in those studies typically involved preparing papers with a final examination consisting of essay questions about course material. The validity of their conclusions is uncertain for quantitative engineering courses where the type of problems given on examinations are similar to those worked during the semester as homework.

DESCRIPTION OF STUDY

In this study, the performance of students from two sections of a required introductory class on fluid mechanics was evaluated using either a series of either unannounced quizzes or two announced midterms. The lectures for the two sections were virtually identical. The quiz problems were similar to the problems given on the midterms, as was the total number of quiz problems given during the semester. The total time spend on evaluation during the semester was the same for the two sections. The two sections were given a common final examination at the end of the semester. The students were also given an unannounced follow-up examination four months later in a subsequent class.

The section receiving the unannounced quizzes had 14 students completing the semester and the section receiving the announced midterms had 19 students. The cumulative GPA of the students entering the class was 2.91 for the quiz section and 2.80 for the midterm section. The cumulative probability distribution functions for the entry GPAs for the two sections are plotted in Fig. 1.

To determine whether the higher GPA of the students in the quiz section was statistically significant, the t' test was conducted. This test is similar to the t test except that it allows the standard deviation of the GPAs in the class to be used instead of the standard deviation of the underlying student population from which the students came. For this test, the variable t' is defined as7

\[ t' = \frac{M_q - M_m}{\sqrt{\frac{S_q^2}{n_q} + \frac{S_m^2}{n_m}}} \]  

(1)
and the associated degrees of freedom are
\[ \nu = \frac{\left( \frac{S_q^2}{n_q} + \frac{S_m^2}{n_m} \right)}{\left( \frac{S_q^2}{n_q^2} + \frac{S_m^2}{n_m^2} \right)} - 2 , \]

where \( M_q \) and \( M_m \) are the calculated mean GPAs of the quiz and midterm sections, respectively, \( S_q \) and \( S_m \) are the calculated standard deviations of the two sections, and \( n_q \) and \( n_m \) are the number of students in each section. The calculated standard deviations of the two sections are found through
\[ S_i^2 = \frac{\sum_{j=1}^{n_i} (x_{ij} - M_i)^2}{n_i - 1} , \]

where \( x_{ij} \) are the individual GPAs of the students in the \( i \)th section.

For the grades given in Fig. 1, the \( t' \) statistic is equal to 0.094 and the degrees of freedom are 31. For the hypothesis that the underlying mean GPAs of the quiz and midterm section are equal, e.g., the students from each section represent a random selection from the same underlying student population, the standard \( t \)-test tables can be used. In particular, the hypotheses that the underlying mean GPAs for the two sections are equal is accepted if
\[ |t| < t_{\alpha/2, \nu} , \]

where \( t_{\alpha/2, \nu} \) is obtained from the \( t \) distribution tables at a level of significance (probability) of \( \alpha \) and at \( \nu \) degrees of freedom. At a level of significance of 95\% (a 95\% probability of being correct) and 31 degrees of freedom, \( t_{0.025,31} \) is equal to 2.04. This value is clearly greater than the calculated \( t' \) statistic of 0.094, so the hypothesis that the underlying mean GPAs of the quiz and midterm section are equal is accepted.

The next step is to determine whether the standard deviations of the two sections are equal. For this test, the \( F \) distribution is used.\(^7\) The \( F \) test statistic is defined as
\[ F = \frac{S_q^2}{S_m^2} . \]

For the two sections, this statistic is equal to 1.113. The hypothesis that the standard deviations of the underlying populations of the two sections are equal is accepted if
\[ \frac{1}{F_{\alpha/2, n_m-1, n_q-1}} \leq F \leq F_{\alpha/2, n_q-1, n_m-1} . \]
For a 95% level of significance, $F_{0.025;19;14}$ is 2.86 and $F_{0.025;14;19}$ is 2.65. Substituting these values into Eq. 6 shows that the inequality is valid. Thus, the hypothesis that the standard deviations of the two sections are equal is accepted.

In the above analyses, it has been shown that both the means and standard deviations of the underlying populations for the two sections are statistically equal. This indicates that the students in each section can be assumed to be randomly drawn from the same underlying population and the slightly higher GPAs of the quiz section was not statistically significant. Thus, it is concluded that the academic abilities of the two sections were equal.

RESULTS OF STUDY

The average score on the common final examination was 69% for the quiz section and 75% for the midterm section. The cumulative probability distribution function for the final examination grades is shown in Fig. 2. For these distributions, both the $t'$ and $F$ statistics were calculated and the hypotheses that the means and standard deviations of the grades of the two sections are equal were tested. The value of the $t'$ statistic was 0.24 and the $F$ statistic was 0.96. These values indicate acceptance of the hypotheses that the means and standard deviations are equal at the 95% confidence level. From this analysis, it was concluded that there was no statistically significant difference in performance on the common final for the two sections. This analysis, however, did not test the original hypothesis that unannounced quizzes lead to an improved long-term retention of the course material.

The average score on the follow-up examination was 49% for the quiz section and 55% for the midterm section. The cumulative probability distribution function for the final examination grades is shown in Fig. 3. For these distributions, both the $t'$ and $F$ statistics were calculated and the hypotheses that the means and standard deviations of the grades of the two sections are equal were tested. The value of the $t'$ statistic was 0.26 and the $F$ statistic was 1.39. These values indicate acceptance of the hypotheses that the means and standard deviations are equal at the 95% confidence level. For the follow-up examination, there were eleven students from the midterm section and seven students from the quiz section. From this analysis, it was concluded that there was no statistically significant difference in performance on the follow-up examination for the two sections. Thus, the original hypothesis that unannounced quizzes lead to better long-term retention than announced midterms is not supported.

DISCUSSION OF RESULTS

Kazmierski claimed that when material is learned at low levels in Bloom's taxonomy it is placed in short-term memory and that material learned at high levels in the taxonomy is placed in long-term memory. Since "cramming" generally involves learning at lower taxonomy levels, material learned by that method is not expected to be retained. One explanation for the similarity in long-term retention for the two sections that was observed in the present study is that the course was structured so that learning occurred at the higher levels of Bloom's taxonomy for both sections. Thus, sufficient material was placed in long-term memory so that the different testing methods had no observable impact. It is noted that the lectures and discussions of both sections were structured to expose students to learning at a variety of levels in Bloom's taxonomy.
Haynie\textsuperscript{10} reported that there was no statistically significant difference in long-term retention for students given in-class examinations or take-home examinations as long as the examinations tested at the higher levels in Bloom's taxonomy. This suggests that properly designed tests can minimize any effects of "cramming" on long-term retention. The quizzes and examinations given in the present study were designed to test understanding at a range of levels of Bloom's taxonomy, including the higher levels.

The homework assigned during the semester to both sections was also designed to facilitate student learning at all levels of Bloom's taxonomy. The grades received on homework assignments by the students in the present study were also analyzed with both the t' and F tests. No statistically significant difference in average homework scores between the two sections was observed. This result is consistent with the results of the previously mentioned study of Conway et al.\textsuperscript{5} that indicated a correlation may exist between grades obtained on course work and long-term retention. The similarity in course work grades observed in the present study would indicate no difference in long-term retention.

One of the limitations to conclusions obtained in this study was the small sample sizes available. This prevented the evaluation of any marginal differences in long-term retention of material. Nevertheless, the wide variation in long-term retention for both sections and the lack of a statistically significant difference between them suggests that varying only the testing method will not result in a significant change in long-term retention. Faculty wishing to maximize long-term retention should first ensure that the course is structured to educate at all levels of Bloom's taxonomy.
STUDENT RESPONSE TO STUDY

The students in both sections were advised of the experiment at the start of the semester. Toward the end of the semester, they were given the opportunity to comment on the experiment. The students were given a questionnaire with the following statement for the hypotheses:

"Having a series of unannounced quizzes instead of a few announced midterms encourages students to study the material over a longer period of time, allowing more material to be placed in long-term memory and retained for a longer time after the end of the semester."

The students were asked if they agreed with the hypothesis and which type of examination method they would prefer in future classes, if they had a choice. The results of this questionnaire are summarized in Table 1. There was a strong correlation between the student responses and which section they were in. Although most students in both the quiz and midterm sections agreed with the hypothesis, only a slim majority of those in the midterm section agreed. Interestingly, a majority of students in each section favored the type of examinations given in their section. Students given quizzes said they would prefer quizzes in future classes and vice versa. Because the sections were selected at random after the start of the semester and students were not allowed to change sections after the selection was made, the students were unable to pre-select the type of examinations they wanted.

Following the announcement to the students that long-term retention was not affected by how they were evaluated during the semester, the students were again questioned about their preference. As shown in Table 2, all of the students who were in the announced midterm section preferred to have announced midterms in future classes. For students that were in the quiz section, however, 44% preferred to continue with quizzes and 55% preferred midterms in future classes. There was no correlation with GPA with this preference. From this survey, the following hypothesis was formed: students with little or no experience with unannounced quizzes tend to prefer announced midterms, while students who have experienced unannounced quizzes are split roughly in half on which they prefer.

CONCLUSIONS

A hypothesis that having a series of unannounced quizzes instead of a few announced midterms would change student study habits and place more information in long-term memory and result in improved long-term retention of the material was tested. It was shown that at a 95% confidence level, the hypothesis was not valid for this quantitative engineering science course. Long-term student performance for a section having unannounced quizzes and for a section having announced midterms was not statistically different. When learning and testing occurs at high levels in Bloom's taxonomy, substantial material is placed in long-term memory and retention is independent of testing method, regardless of how much "cramming" occurs. Students not having experienced unannounced quizzes prefer announced midterms, while students who have experienced unannounced quizzes tend to be evenly split in preference between announced midterms and unannounced quizzes in future classes.
REFERENCES


BIOGRAPHICAL INFORMATION

JOHN C. REIS earned his Ph.D. in Mechanical Engineering from Stanford University. He has worked for Chevron, Los Alamos National Laboratory, Lawrence Livermore National Laboratory, as a private consultant, and currently teaches at Embry-Riddle Aeronautical University. In 1995, he received the Ralph R. Teetor award from the Society of Automotive Engineers for excellence in engineering education.
Table 1  Results of Student Opinion of the Hypothesis

<table>
<thead>
<tr>
<th></th>
<th>Agree with Hypothesis</th>
<th>Disagree with Hypotheses</th>
<th>No Opinion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quiz Section</td>
<td>87.5%</td>
<td>12.5%</td>
<td>0%</td>
</tr>
<tr>
<td>Midterm</td>
<td>55%</td>
<td>40%</td>
<td>5%</td>
</tr>
</tbody>
</table>

Table 2  Student Preference Knowing Hypothesis is Not True

<table>
<thead>
<tr>
<th></th>
<th>Prefer Quizzes</th>
<th>Prefer Midterms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quiz Section</td>
<td>62.5%</td>
<td>25%</td>
</tr>
<tr>
<td>Midterm</td>
<td>10%</td>
<td>75%</td>
</tr>
</tbody>
</table>

Figure 1. Cumulative Probability Distribution for Entry GPA

1.454.8
Final Course Grade

Figure 2. Cumulative Probability Distribution for Final Grade

Follow-up Grade

Figure 3. Cumulative Probability Distribution for Follow-up Examination