The Effectiveness of Video Quizzes in a Flipped Class

Conference Paper · February 2015

DOI: 10.1145/2676723.2677302

2 authors:

Mark C. Lewis
Trinity University
97 PUBLICATIONS **509** CITATIONS
SEE PROFILE

Lisa Lacher
University of Houston - Clear Lake
11 PUBLICATIONS **82** CITATIONS
SEE PROFILE

The Effectiveness of Video Quizzes in a Flipped Class

Lisa L. Lacher
School of Science and Computer Engineering
University of Houston - Clear Lake
2700 Bay Area Blvd.
Houston, TX, USA
Lacher@uhcl.edu

Mark C. Lewis
Department of Computer Science
Trinity University
1 Trinity Place
San Antonio, TX, USA
mlewis@trinity.edu

ABSTRACT

The use of online videos and flipped classrooms is a natural fit for many CS courses, especially in the introductory sequence, and the use of these techniques is growing. However, the benefits of these techniques depend heavily on student preparation before class meetings. This paper presents tests of the effectiveness of basic multiple choice quizzes as gateway checks in a flipped classroom by looking at the impact they have on student grades. Programming aptitude and learning approaches were considered as additional factors. Ignoring aptitude and learning approaches, statistical analysis of the student grades did not support our hypothesis that the gate-check video quizzes would be beneficial. Nor did it support our hypothesis that the benefit of quizzes would be greatest for students who had lower aptitudes entering the class. However, there was some evidence that it might have helped raise the grades of students who have a propensity toward surface learning.

Categories and Subject Descriptors

K.3.2 [**Computer and Information Science Education**]: Computer Science Education

General Terms

Measurement

Keywords

flipped classroom; CS1; video quizzes

1. INTRODUCTION

The concept of a flipped classroom, where students are introduced to material outside of class and class time is used to apply that material, has gained significant momentum in the past several years with the advent of ubiquitous online videos and simple playback. The idea is that students watch the videos, which contain normal lecture material, before coming to class then spend their time in class doing exercises or assignments. This approach has been embraced in Computer Science, as it allows students to spend more of their valuable class time doing hands-on activities instead of sitting passively absorbing information [11, 19, 10, 1]. In the ideal implementation, students figure out in class what things they don't fully understand, and are able to ask the instructor for clarification at that time.

Of course, the effectiveness of this approach depends significantly on students showing up to class prepared. Earlier studies have shown mixed results based on how well students prepare. Students who fail to watch the videos in advance will generally be completely lost during the class time, which one would expect to lead to inferior learning outcomes. If a sufficient fraction of the class fails to prepare, the instructor can be forced to revert to a standard lecture format to make sure that students are getting the most important material. This can be hard to prevent as often both the students and the instructor have a long history with the traditional lecture format, and are comfortable using it.

For this reason, it would be beneficial for instructors who use a flipped classroom approach to find a mechanism that will increase the odds that students will show up to class having done the recommended preparation work. The goal of this study was to see if having an element of the course grade come from online, multiple-choice quizzes over the video material, taken prior to each class meeting, was effective in this regard. Early results of this study, including grades up to the course midterm were presented in [14]. The primary hypothesis in this earlier paper was that pre-class video quizzes (gate-check quizzes) would help students earn better grades, which we assumed would indicated that better grades would correlate with better prepared students. The midterm grade results did not support this hypothesis and we speculated that the quizzes failed because they did not facilitate deep learning. The secondary hypothesis in the earlier paper was that the pre-class video quizzes would help students with lower computer programming aptitudes perform better in the course. We found that the midterm grade results did not support this hypothesis either. In fact, the results indicated that it was the higher aptitude students who benefited most from having the video quizzes. This work adds additional grades to the analysis, and also takes into account student learning approaches, which were not considered previously.

SIGCSE'15, March 4–7, 2015, Kansas City, MO, USA.

http://dx.doi.org/10.1145/2676723.2677302.

2. EXPERIMENTAL DESIGN

These classes were taught using [16]. A full set of video lectures, done as screencasts, have been posted publicly on YouTube [15] for any instructors to use in flipping their classrooms. Students were required to watch these videos in advance of covering each chapter in class. This study was done in the spring of 2014 with three sections of CS1 taught at Trinity University by two different faculty. To determine the effectiveness of the video quizzes, some sections (experimental group) were given quizzes that counted as 10% of their grade in the course whereas other sections did not take the video quizzes and therefore could not be graded on them (control group).

The students were also given three surveys at the beginning of the semester.

- A background information survey to collect data such as their major and their previous background with programming.
- The University of Kent Programming Aptitude Test [18], to measure the students aptitude for programming upon entering the course.
- A learning approaches survey from Royal Free & University College Medical School, London, UK to measure if students had a tendency toward surface, deep, or achievement learning.

We consider eight written assessments given during the course of the semester: six written quizzes, the midterm, and the final exam. Some of the students were required to take gate-check video quizzes. These quizzes were done online before class began and included a mix of multiple choice, true/false, and matching questions that were automatically graded. Students were given multiple attempts at the quizzes. After some experimentation early in the semester, we settled on giving students two attempts to get correct answers for the quizzes.

As the name, gate-check video quiz, implies, these quizzes were created to be simple gate-checks, not assessment of deep learning or skills that the students acquired from watching the videos. The questions were generally simple to answer if students had actually watched the videos. They were also presented in the order that they appeared in the videos. The entire goal of these video quizzes was to provide a strong incentive for students to watch the videos, not to stress items that were important, or to facilitate their learning of the material in the videos.

2.1 Programming Aptitude Test

The programming aptitude test developed at the University of Kent [18] includes twenty six questions that deal with numerical problem solving, logical reasoning, attention to detail, pattern recognition, and the ability to follow complex procedures. It does not involve code/programming topics directly. This makes it appropriate for introductory CS students, as it tests them on skills that are significant for the task of programming without actually testing their knowledge of programming. Students were given forty minutes to complete this test.

2.2 Approaches to Learning and Studying

Students approach studying in different ways, and the way a student approaches a learning task should have an effect on the quality of their learning. One learning theory approach draws a distinction between deep and surface learners [17].

Students who use a deep learning approach intend to look for meaning in study materials. They will sift out general principles and concepts and then relate these concepts to prior knowledge or everyday life by closely examining the material's content. This type of student generally has an interest in the subject; and this interest may be derived from the subject being vocationally relevant, or the student may just want to personally understand the material.

The main desire of students who use a surface learning approach is to only meet a task's requirements such as passing an exam. They want to limit what they have to learn to the bare essentials and they have little real interest in the content. The most common strategy adopted by students using surface learning is rote learning or memorizing the study materials instead of understanding them [17]. Some researchers found that the difference between deep learners and surface learners could be explained by an active or passive attitude to learning [5]. Distinctions between deep and surface learning are also made, where motivation increases deep-level processing and anxiety increases surface-level processing [8]. Surface learners fear failure.

Besides the distinction between the deep and surface approach to learning, further research added a third approach to learning: the achieving (or strategic) approach [4]. This is an achieving orientation, which is about maximizing performance while optimizing efforts to achieve it. The student does this by using whatever technique (deep or surface) that will produce the highest grade, and organizing their time in an effective way to fulfill course requirements [20].

Learning approaches have been reported to have an impact on the outcome of learning; high academic achievement has been positively related to a deep or strategic approach and negatively related to a surface approach [6, 7]. This relationship occurs provided that the assessment requires high levels of understanding from the students [2]. It is generally assumed that the deep learning approach results in higher quality learning outcomes than the surface learning approach [13, 12]. In other words, students who employ more surface learning strategies will not perform as well as students who employ more deep learning strategies. It is also interesting to note, that a deep learning approach has a positive impact on self-directed lifelong learning [2].

2.2.1 *Measuring Learning Approaches*

The Study Process Questionnaire (SPQ) is a widely used measure of learning approach. Students' scores on the scales of the SPQ are most naturally interpreted as measuring their predispositions to adopt different approaches to learning. The Biggs Study Process Questionnaire (SPQ) is a diagnostic tool to identify the learning approaches of the students. The SPQ, a forty two item questionnaire, provides feedback on the learning approaches in the three dimensions, namely, the surface, deep and achieving dimensions which are further broken down into motives and strategies used. Fox et. al [9] created a shortened eighteen item version of the SPQ and conducted a longitudinal study to measure the stability of the instrument over five to seven years. They concluded that the shortened eighteen item SPQ has proven reliable and stable. Because the eighteen item version was easier to administer to students, we choose to use the shortened SPQ developed by Fox et. al. in our study [9].

2.3 Hypotheses

We had three hypotheses related to requiring the gate-check video quizzes as a means of enforcing that students prepare before coming to class.

1. The students taking the gate-check video quizzes have higher grades on the assessments than those who do not have to take the video quizzes.

2. Students with lower aptitude in programming benefit more from the gate-check video quizzes than those who do not.

3. Students who have a propensity for shallow learning benefit more from the gate-check video quizzes than those who have a propensity for deep learning.

The logic behind the first hypothesis is the reason why we considered making the gate-check video quizzes in the first place. Flipped classrooms only work well when students come to class prepared, and having students participate well in class should improve their grades. The second hypothesis is based on the belief that high aptitude students generally do well and probably would not benefit much from the quizzes, while low aptitude students really need the preparation that the video quizzes provide. The third hypothesis is similar to the second one. Our assumption is that those with a propensity for shallow learning will typically do poorly in CS1 compared to those with a propensity for deep learning. As such, shallow learners have the most to gain. We also believed that the videos, which show live coding and the logic that goes into creating code, would lead students who normally tend toward shallow learning to actually make deeper connections in the material.

2.4 Participants

There were forty seven undergraduate students enrolled in the three sections of Principles of Computer Science I course at Trinity University during the spring semester of 2014. Thirty-two of these students participated in the study and completed quizzes and tests. These sections were taught by two different professors and each section included between fifteen and eighteen students. All students were urged to watch the videos and/or read the book before attending class. Two of the sections, one taught by each instructor, included the gate-check video quizzes as part of their grades.

2.5 Artifacts

The courses used identical written quizzes and tests. The quizzes were graded by a student grader using a rubric developed by the instructors. The exams were graded by the instructors, with each instructor grading all of a particular question, so that each question was graded uniformly across all sections.

Students also completed coding assignments, but those are not included in this study because the results tend to be Boolean in nature. Students who put in the time and effort to complete the assignments get full or nearly full marks, while those who do not get few, if any, points. As we don't have any method of determining how much time and effort students put into the coding assignments, assignment grades to not provide a useful way to analyze the impact of the gate-check video quizzes.

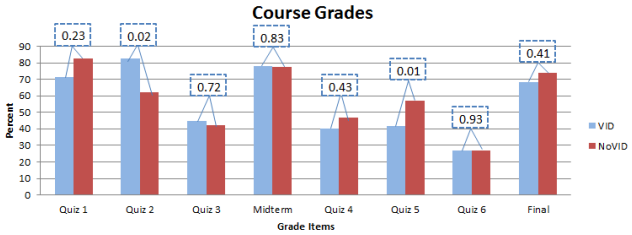

Figure 1: This figure compares the quiz and test grade averages for students who did (VID) and did not (NoVID) have to take the gate-check video quizzes. The differences are generally small. The number above each pair of grades is the t-test p value. Only the p value for quizzes 2 and 5 are below the $\alpha = 0.5$ criteria for significance.

3. ANALYSIS AND RESULTS

To check the first hypothesis, we look at the grades for the different assessments, comparing results from students who had to do the gate-check video quizzes to those who did not. The results of such a comparison are shown in Figure 1. This figure makes it clear that there were not large differences between the two groups. An independent samples t-test was performed on each of these different grades and only quizzes 2 and 5 have a statistically significant difference of $p < \alpha$ for $\alpha = 0.05$. In the case of quiz 2, the students who were taking the gate-check video quizzes scored higher, while the opposite was true for quiz 5.

It is also interesting to note that while the results were very mixed in the first half of the semester, up to the midterm, the group that did not use the gate-check video quizzes had higher grades for all the quizzes after the midterm and the final exam. Granted, the difference was only statistically significant for quiz 5.

These results lead us to conclude that our first hypothesis was incorrect. The gate-check video quizzes did not have a significant positive impact on student grades. Fortunately, they don't have a significant negative impact either.

To test the second hypothesis, we divide the students into three categories, as prescribed by the creators of the programming aptitude test. Students who correctly answered nineteen or more of the twenty-six questions are said to have a high aptitude. Students who correctly answered between thirteen and eighteen questions have an average aptitude. Those who correctly answered fewer than thirteen questions have a low aptitude.

Figure 2 shows a weighted average of the quizzes and tests for each of these aptitude groups for all the students. The average shown is calculated as a weighted average of the student's quiz and exam averages, with the quizzes counting as one third and the exams counting as two thirds. The first thing to note about this is that in both groups, higher aptitudes are correlated with higher grades. It should also be noted that the group that did not take the gate-check video quizzes did not include any low aptitude students. This is the first evidence that our student distribution was not very uniform.

The relative performance of the students in each of the aptitude levels across the VID and NoVID groups is made clearer if we bin the students together by whether they are

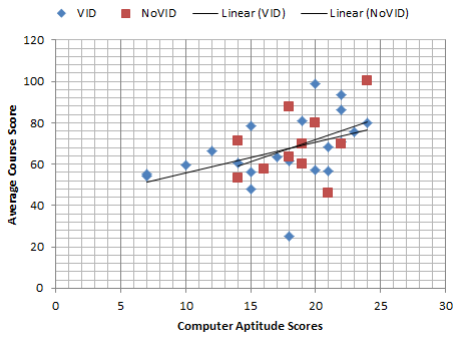

Figure 2: This figure shows weighted course averages as a function of student aptitudes. For both the VID and NoVID groups, there is a positive correlation between aptitude and grades, but there is no clear separation between the VID and NoVID groups.

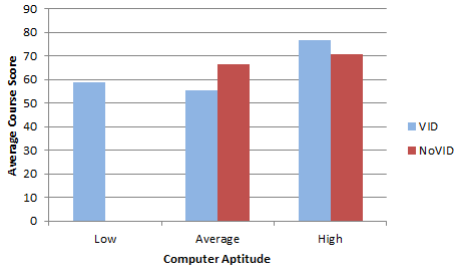

Figure 3: This figure shows weighted course averages binned by student aptitudes. There were no low aptitude students in the NoVID group. The differences in the average and high aptitude groups are not statistically significant according to a t-test.

high, average, or low aptitude. This produces Figure 3. Standard t-tests show that neither the difference in the average aptitude groups, nor the high aptitude groups are statistically significant. For this reason, we have to conclude that our second hypothesis was also false, and that the gate-check quizzes do not have a significant impact on the non-high aptitude students. Unfortunately, the lack of low aptitude students in the NoVID group makes it harder to draw additional conclusions. The average grade of the low aptitude students in the VID group is higher than those in the average aptitude group, but this is largely because of a single outlier in the average group with a very low grade. Reasonable extrapolations for the performance of the NoVID low aptitude students would put them in a range that is consistent with those in the VID group.

To test the third hypothesis, we first looked at the distribution of learning approaches between the students who did and did not take the gate-check video quizzes. This is shown in Figure 4. The p values shown above each style indicate that the differences for the surface and deep approaches are statistically significant, while that for achievement learners is not.

According to [3], students who tend toward deep learning do better in CS1. Based on this, and the skew in our sample,

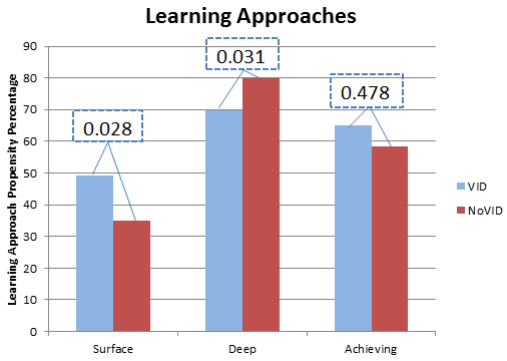

Figure 4: This figure shows how students answered the learning approaches quizzes. For each approach there were several questions which students could answer as tending toward or away from that approach. The vertical axis shows what percentage of the questions were answered in the style of that approach across all students in both the VID and NoVID samples. Our distribution of students was skewed in a statistically significant way such that the NoVID students had a greater average tendency toward deep learning, while the VID students had a greater average tendency toward surface learning.

we would expect that the students who were not taking the gate-check video quizzes should have higher grades on average than those who were taking the video quizzes. However, as was shown in Figure 1, this is not the case. Instead, the two groups had grades that were statistically indistinguishable. This provides some evidence that the third hypothesis might actually be correct, and that the gate-check video quizzes helped to boost the grades of the students who had a propensity for surface learning sufficiently to bring them in line with those who had a propensity for deep learning. Unfortunately, we have not been able to pull apart the effects of the gate-check video quizzes and the learning approaches in our data, so this is weak support for the hypothesis at best.

4. CONCLUSIONS

The most obvious conclusion from this work is that simple, gate-check video quizzes are not sufficient to improve the grades of students in a flipped CS1 classroom. Granted, that conclusion comes with the proviso that we were looking at a small sample size, and the enrollments in the studied class sections had some distinct skew that might have favored the group that was not required to take the video quizzes. However, even with this proviso, it seems reasonable that this is a valid conclusion because of the limited nature of the gate-check video quizzes.

The quiz style that was used for this work was a surface learning tool, not something that really prompted students to think more deeply, make connections, or question their assumptions. It was unlikely that students who had watched the videos only somewhat attentively would find that they couldn't answer the questions and need to go back to re-

watch sections, or go to the book for additional assistance.[1] It is quite possible that some form of pre-class assessment could be found that would force students to take these steps, and which might actually produce better results. For that reason, we are looking into using a pre-class assessment that asks students to answer questions that involve filling in gaps in code segments or writing functions to do very simple calculations instead of using quizzes with multiple choice questions. We expect that such an approach would actually facilitate deeper learning by students, and would therefore be more effective at improving their learning.

Another factor to consider is that this work looks only at outcomes in terms of student grades. What we are truly interested in is student learning, and their ability to retain and apply the knowledge from the class. Grades are a convenient numerical proxy for learning, but they are certainly not a perfect metric. Even though the grades did not reflect an improvement due to the gate-check video quizzes here, one of the instructors noted anecdotally that the students were more uniformly prepared when required to take the gate-check video quizzes than they had been in previous semester when they were not.

Over the course of the semester, it is inevitable that students become more or less busy at various times. When their other classes begin to give them a lot of work, things like video watching are often some of the first things to suffer. While these students might be able to cram later to catch up on things for quizzes and tests, the failure to watch videos impacts the flow of the course as there can be days when many students clearly show up unprepared. For one instructor in this study, the gate-check video quizzes did seem to come close to eliminating this behavior. Why that didn't improve grades for students is a separate issue that would be worth exploring.

5. REFERENCES

[1] A. Amresh, A. R. Carberry, and J. Femiani. Evaluating the effectiveness of flipped classrooms for teaching cs1. In *Frontiers in Education Conference, 2013 IEEE*, pages 733–735. IEEE, 2013.

[2] B. Anderson, S. Lee, M. Simpson, and S. Stein. Study orchestrations in distance learning: Identifying dissonance and its implications for distance educators. *The International Review of Research in Open and Distance Learning*, 12(5), 2011.

[3] O. Bälter, M. Cleveland-Innes, K. Pettersson, M. Scheja, and M. Svedin. Student approaches to learning in relation to online course completion. *Canadian Journal of Higher Education*, 43(3):1–18, 2013.

[4] J. B. BIGGS. Dimensions of study behaviour: Another look at ati. *British Journal of Educational Psychology*, 46(1):68–80, 1976.

[5] L. O. Dahlgren and F. Marton. Students' conceptions of subject matter: An aspect of learning and teaching in higher education. *Studies in Higher Education*, 3(1):25–35, 1978.

[1]In a survey of students reported in the earlier paper, students do self-report that a majority of them re-watched parts of the video lectures that were unclear to them. However, we were unable to determine if that was because of the quizzes or their reaction to not understanding the material.

[6] A. Diseth. Personality and approaches to learning as predictors of academic achievement. *European Journal of Personality*, 17(2):143–155, 2003.

[7] A. DISETH and A. MARTINSEN. Approaches to learning, cognitive style, and motives as predictors of academic achievement. *Educational Psychology*, 23(2):195–207, 2003.

[8] N. Entwistle. Strategies of learning and studying: Recent research findings. *British Journal of Educational Studies*, 25(3):225–238, 1977.

[9] R. A. Fox, I. McManus, and B. C. Winder. The shortened study process questionnaire: An investigation of its structure and longitudinal stability using confirmatory factor analysis. *British Journal of Educational Psychology*, 71(4):511–530, 2001.

[10] G. Gannod, J. Burge, and M. Helmick. Using the inverted classroom to teach software engineering. In *Software Engineering, 2008. ICSE '08. ACM/IEEE 30th International Conference on*, pages 777–786, May 2008.

[11] E. F. Gehringer and B. W. Peddycord III. The inverted-lecture model: a case study in computer architecture. In *Proceeding of the 44th ACM technical symposium on Computer science education*, pages 489–494. ACM, 2013.

[12] D. Gijbels, G. Van de Watering, F. Dochy, and P. Van den Bossche. The relationship between students' approaches to learning and the assessment of learning outcomes. *European Journal of Psychology of Education*, 20(4):327–341, 2005.

[13] T. Hau, S. Williams, and P. Hoi. Using the biggs' study process questionnaire (spq) as a diagnostic tool to identify "at-risk" students : A preliminary study, 2004, Retrieved September 5, 2014.

[14] L. L. Lacher and M. C. Lewis. The value of video quizzes in a computer science flipped classroom: An empirical study. In *Proc. Int. Conf. on Frontiers in Education: Computer Science and Computer Engineering (FECS'14)*, pages 94–100, Las Vegas, USA, July 2014. CSREA.

[15] M. C. Lewis. Video playlist: Introduction to the art of programming using scala (part i). `https://www.youtube.com/playlist?list=PL0B0820169DCF0AD2`.

[16] M. C. Lewis. *Introduction to the art of programming using Scala*. CRC Press, 2012.

[17] F. Marton, R. Säljö, et al. Approaches to learning. *The experience of learning*, 2:39–58, 1984.

[18] U. of Kent. Computer programming aptitude test. `https://www.kent.ac.uk/careers/tests/computer-test.htm`, April 2014.

[19] A. Sureka, M. Gupta, D. Sarkar, and V. Chaudhary. A case-study on teaching undergraduate-level software engineering course using inverted-classroom, large-group, real-client and studio-based instruction model. *CoRR*, pages –1–1, 2013.

[20] H. Tait and N. Entwistle. Identifying students at risk through ineffective study strategies. *Higher Education*, 31(1):97–116, 1996.