THE IMPACT OF INDIVIDUAL ONLINE TESTS IN ADDITION TO GROUP ASSIGNMENTS ON STUDENT LEARNING

Ronald Huisman and Helen E. Reedijk
Erasmus University Rotterdam
The Netherlands

Abstract

In this paper we examine whether a specific setup of online tests improves students' learning. To do so, we study the results of students that followed the Mathematics course within the Bachelor 1 program of the Erasmus School of Economics between September and October 2011. During this course, students worked in groups completing exercises assisted by teachers. On top of that, the students had the opportunity to participate in weekly tests provided on their digital learning platform. These tests consisted of exercises with which they could test their individual knowledge. After doing the test, the students received automated individual feedback and could redo the test in order to apply the feedback directly. We found, while controlling for the students' entry level of knowledge measured through a diagnostic test made at the beginning of the course, that students who actively took the tests obtained a significant higher probability of passing the exam and obtained a significant higher grade for the course.

Introduction

Starting in September 2011, 630 students participated in a seven weeks mathematics course as a part of the Bachelor 1 program 'Economics and Business' at the Erasmus School of Economics in Rotterdam. During the course, two teaching methods were applied to increase student time on task. First, the students worked in groups on assignments during weekly classroom sessions. Second, weekly individual online tests were offered as a formative testing tool to provide students individual feedback on their actual level of knowledge in relation to the desired level of course content knowledge. In this paper, we examine whether the weekly online individual tests in addition to the weekly group assignments improved student learning. We think that this is worth examining as there is mixed evidence on the effect of frequent testing on student learning in the literature.

Some studies show that the impact of online testing is questionable. For instance, Steenhuis, Grinder, and de Bruijn (2009) analyzed the relationship between online quiz performance and the final course grade in an introductory operations management class. They found evidence for a positive relation, but the authors argued that both grades could reflect students' ability and that the relationship found is not a good indicator of the added value of online quizzes. Another result of their study also raises questions about the use of online quizzes. A comparison was made between a teaching period during which online quizzes were used and a teaching period without quizzes. The course grades were not significantly higher in the period with online quizzes. In another study, Haberyan (2003) also did not find that weekly quizzes are an effective way to improve students' performance. Students in a general biology course had to complete quizzes in the beginning of the lecture. Exam grades were not significantly higher after introducing these online quizzes.

Other studies show that online testing improves student learning. Kibble (2007) studied the effect of the use of unsupervised online quizzes in a Medical Physiology course. From 2004 till 2006 formative multiple-choice quizzes were used in five different ways, varying in available course credit per quiz and criteria for earning those credits. It was found that students were more willing to participate in the online quizzes when the quizzes were graded and the grade counted as a certain part of the final course grade. Students who did chose to participate in the online quizzes generally performed better on the final exam than students who did not participate in the online quizzes. Charman and Elmes (1998) found evidence that students' performance significantly increased after introducing weekly tests in a first-year course geographical data analysis. Final exam performance was compared before and after the introduction of weekly tests. The results showed a small increase in average exam grades. The tests were especially beneficial to students who were struggling with the course content and on the borderline of passing or failing the course. Another study that found a positive effect of formative online quizzes on students' performance on the final exam is the study of Cluskey, Hodges, and Smith (2006). Web-based quizzes for each chapter from the course book were introduced in an introductory Financial Accounting class to make students study continuously. The students were allowed to take all tests three times, and the highest grade counted as a part of their final grade. When comparing the course with online quizzes and the old-style course without online quizzes, the authors found increased course pass rates, increased average exam grades and lower drop rates.

We examine the effect of online quizzes in a specific setting. The mathematics course that we examine offers two methods to activate students: weekly group assignments and the individual online tests. The setup of the online tests in the course is comparable to that of Angus and Watson (2009) and Kibble (2007). In those studies, online quizzes were designed such that two attempts were allowed and the highest grade of the two tests accounted for the final course grade. And, as in Daniel and Broida (2004), the mathematics course that we examined put time limits on the tests and randomly selecting questions out of a database in order to reduce inappropriate use. From this particular design of online tests, we know that it improves student learning. Angus and Watson investigated the link between online quizzes and student learning by testing exposure to the online learning environment and exam performance. In a first-year business mathematics course, the authors found that exposure to the online testing system significantly improved student learning, measured by a final exam grade. Students who completed a higher than average amount of online tests achieved a higher final exam grade. Varsavsky (2004) studied the effect of introducing online weekly quizzes with different question types in a first-year calculus course. The purpose of introducing these weekly tests was to provide students with feedback and to encourage learning. The quizzes were found to be an effective tool in students' learning process and introducing these weekly quizzes helped students to perform better on the final course exam. Griffin and Gudlaugsdottir (2006) studied the effect of a project that was introduced in mathematics and operations research courses in which random online quizzes with unlimited attempts were offered to students. The quizzes were offered as formative learning tools and had a positive effect on students' performance and learning behavior. An additional advantage the authors noticed was that students asked for teachers help when needed more frequently due to the fact students were early in the study process aware of existing difficulties with certain topics.

Brothen and Wambach (2001) made a distinction between two strategies in which students could use formative quizzes: a quiz-to-learn strategy and a prepare-gather feedback-restudy strategy. The latter was found to be a more successful strategy leading to improved exam performance compared to using the quizzes to learn the course material, which was found to be an ineffective study technique. The online tests in the mathematics course that we examine were designed following the prepare-gather feedback-restudy strategy.

Our contribution to the previous literature is that we examine the effects of the individual online tests in addition to the weekly group assignments. According to the previous studies, we expect that the specific online tests increase student learning, but we don't know whether these tests also add value in a structure in which student learning is already activated through weekly group work. Secondly, keeping in mind the remarks made by Steenhuis et al. (2009), we contribute by testing the robustness of the result, controlling for students' ability reflected by their entry level of knowledge. We therefore use information from the students' secondary school type of mathematics they followed and the grade they obtained for a diagnostic test that they made on the first day of the course.

In the following section, we describe in detail the design of the course that we examine and then we show the results in the section thereafter.

The Course Structure

The first-year mathematics course that we examine is a compulsory course for all Bachelor 1 students at the Erasmus School of Economics. The duration of the course is seven weeks and it is concluded with a final exam.

On the first day of the course, all students made a computerized test in which they had to answer mathematical questions related to the contents of the whole course. Beforehand, it was communicated to the students that the test was meant as a diagnostic test to provide them and the teachers with insight into their prior knowledge of the course contents. Furthermore, they were told that the score for the test would not be a part of their final grade for the course. The score for the diagnostic test was communicated to all students. The diagnostic test was provided through Maple TA in the student learning platform Blackboard and consisted of eight open quantitative questions with random numbers. Maple TA is a web-based testing tool, in our situation working as a building block within Blackboard, and especially useful for mathematics tests. Students had to finish the test within 45 minutes.

During each of the seven weeks, no general lectures were given. Instead, students had to study chapters in the course book themselves and watch webcasts containing the contents for that specific week. Every week one plenary lecture hour was scheduled by the teacher to answer questions of the students about the topics explained in the webcasts.

The acquired knowledge was applied in weekly practical sessions of two hours, in which the student group was divided into several groups of 25 to 30 students. In these practical sessions the students were provided with a short overview (around 15 minutes) of the topics of that specific week. And after that, the students worked on assignments in groups of four students accompanied by a teaching assistant. These assignments consisted of exercises with open questions covering the relevant topics for that particular week, in which the students were asked to answer the questions and show their way of calculating the answers. The assignments were not part of the final grade for the course. The teaching assistants corrected (but not graded) the assignments to give the students in the group an indication of their knowledge level. For the practical sessions a compulsory attendance of 70% applied.

Furthermore, every week a plenary exercise lecture of two hours was scheduled in which one of the teaching assistants explained problems covering topics of that particular week step by step. The students were asked to prepare themselves for these exercise lectures by doing weekly homework containing quantitative open questions available in their course book. This homework was not compulsory and was not corrected by a teacher or teaching assistant, students were self-responsible for preparing for this lecture, and for participating in this lecture.

Starting in the second week, students could do a digital test provided through Maple TA in Blackboard at the end of every week during the course period (six tests in total). The average score over the six digital tests accounted for 10% of the total grade for the course. Participation in these digital formative tests was not compulsory, and the low attribution in the final grade implies that students could pass the course (by getting a sufficient exam grade) without participating in the weekly online tests. Table 1 shows that more than 90% of the students took the online tests.

Table 1

Participation Grade in Weekly Online Tests

Test	Participation
1	90.63%
2	94.44%
3	93.65%
4	91.59%
5	90.95%
6	90.95%

To enter the tests the students had to log in with their personal username and password and all students' activity was registered in the system. Each test consisted of between six and ten open and multiple-choice quantitative questions. The numbers used in the questions were randomized to reduce the impact of fraud. Students were allowed to take the test twice and the highest grade obtained over the two tests counted as the score for the test. The test grade students could achieve was between 1 and 10. Students had to take the tests between Thursday and Sunday. Each test had to be completed within a period of 60 minutes.

After completion, the students received immediate automated feedback on their score and hints about how to solve the questions that were answered wrongly, for instance, by pointing to the paragraphs in the text book and the webcasts where the theory needed was discussed. As said, the students could do the test again and were thereby offered a way to directly apply the provided feedback and measure their increase in knowledge. Because of the randomization, the second test was similar but not identical to the first test. As the test was provided on the digital learning platform, students could make the tests at any place and time (though within the time window) they preferred. The motivation behind this setting is that both time-on-task and personal feedback enhance student learning (Hattie, 2009).

At any moment during the course period students had access to the results of their online tests, and also the previous test questions, given answers and provided feedback were available any time. The course ended up with a three-hour final exam consisting of a combination of open and multiple-choice questions. The exam grade counted for the remaining 90% of the final grade. The exam grade is measured on a scale between 1 and 10.

Results

The goal of this paper is to examine whether personal online weekly testing in addition to group work improves students learning. In this section, we discuss step by step the methodology and the results.

In order to examine the impact of the online tests on student learning, we compare the average and standard deviation over the final exam grades and the likelihood of passing the exam (we refer to this as the success rate) between two groups of students: a group with students who actively participated in the online tests (we refer to them as the active students) and a group with students who did not (inactive students). We then test the hypotheses that there is no difference between the two groups in terms of average grade, standard deviation of the grades and success rate against alternative hypotheses that are in line with improved student learning. We assume that this occurs when the group of active students show a higher average exam grade, an increased likelihood to pass the final exam and lower standard deviation over the final grades, signaling less randomness in the grades and more cohesion due to better learning.

This methodology implies that we have to distinguish between inactive and active students. We chose to define a student to be active if he or she completed all the six tests and passed all tests by scoring a grade of at least 5.5 for each test[1]. We have two reasons for this definition. First, this definition implies that those students who did not take the tests seriously are seen as inactive students[2]. The second reason is a managerial one. This definition makes it possible to forecast learning results: those students that either did not take the first test or obtain an insufficient grade of that tests are likely to perform worse in the final exam. In this case the online tests turn out to have a positive impact on student learning. Table 2 shows that this definition yields 392 inactive and 238 active students in our sample.

Table 2

Learning Results of Active and Inactive Students

	Inactive	Active	Total
n	392	238	630
grade	6.204	7.303	6.701
	(0.081)	(0.092)	(0.065)
stdev	1.610	1.420	1.621
success	0.508	0.882	0.649
	Standard errors are between parentheses		
Δ grade	(8.948; 0.000)		
Δ success	(11.407; 0.000)		
eq. variance	(1.286; 0.017)		

Table 2 also provides information about the learning results for both groups of students. The inactive students obtained an average exam grade of 6.204 whereas the active students were graded 7.303 on average. The standard deviation of the exam grades was 1.610 for the inactive students and it was lower at 1.420 for the active students. Of the active students, 88.2% passed the exam, whereas only 50.8% of the inactive students passed it. All these differences between the two groups are significant with more than 99% confidence. This can be inferred from Table 2 as follows. The difference between the average grade for active and inactive students is 1.099; the active students scored on average 1.099 higher than the inactive students. The row "Δ grade" in Table 2 shows the result of a one-sided t-test[3] to test the hypothesis that both average grades are equal against the alternative hypothesis that the active students score higher. Table 2 shows that the t-value of this test equals 8.948 and the corresponding p-value is 0.000 indicating that the difference is highly significant. Comparing the success rate (row: Δ success), the % of the students that passed the exam, we observe that the active students had a 37.4% higher success rate. A one-sided z-test[4] yields a z-value of 11.407 and a p-value of 0.000. The difference in success rate is highly significant. The same holds for the difference between the standard deviations of both groups (row:

eq. variance); the standard deviation of the grades of the group with active students is 0.190 lower than the standard deviation of the grades of the inactive students. A one-sided F-test[5] to test whether the variances of both groups are equal has a F-value of 1.286 and a p-value of 0.017 indicating that the standard deviation of the grades of the active students is significantly lower than the standard deviation of the grades of the inactive students. From the results in Table 2 we conclude that those students who actively made the weekly online tests obtained a significantly higher exam grade, were significantly more likely to pass the exam and their grades showed significant less variation. So far, we conclude that active participation in the weekly online tests significantly improved student learning in our sample.

This group of 630 students consists of mostly first-year students, but also some recidivists and pre-master students, who have to finish this course to eliminate their math deficiency. To make the studied group of students more homogeneous we proceed in the remainder of this paper with the subsample of 526 first-year students (i.e., we filter out the recidivists and the pre-master students). Table 3 shows the results of that group consisting of 308 inactive and 218 active students. Again we found higher average exam grades for the active students; active students scored on average 0.816 higher than the inactive students. This difference is highly significant as the Δ grade row indicates[6]. Furthermore, the active students had a 13.1% higher success rate compared to the inactive students; this difference is smaller for this more homogeneous group as it was for the total group of 630 students. This can be explained by the fact that there were some different rules for the recidivists and pre-master students, for example the compulsory attendance rule for the practical sessions did not hold for them.

Both the difference in success rate and standard deviations are significant as well, as is shown in rows Δ success and eq. variance.

Table 3

Learning Results of Active and Inactive Students

	Inactive	Active	Total
n	308	218	526
grade	6.484	7.300	6.822
	(0.096)	(0.095)	(0.070)
stdev	1.679	1.401	1.619
success	0.750	0.881	0.785
	Standard errors are between parentheses		
Δ grade	(6.056; 0.000)		
Δ success	(3.968; 0.000)		
eq. variance	(1.436; 0.002)		

The definition of active and inactive students we used in our study is the result of a discussion about how to define active and inactive students. In our opinion another useful definition can be one based on the average weekly test score obtained over all tests, because the initial definition we use defines students who miss one test but for the rest participated actively as non-active. The alternative definition divides the group of 526 students in 263 inactive students with a below median average weekly test score, and 263 students with an above median average score for the weekly tests.

Table 4 shows the results in case this definition is used. Using this definition the results are comparable with the results when using the previous described definition, again we found significant higher average exam grades and success rates for active students compared to inactive students. We therefore conclude that the choice between one of the two described definitions will not make much of a difference in the results. We will proceed with the initial definition in which an active student is a student that made all weekly tests and obtained a sufficient result for all tests as we think that this definition is attractive from a managerial point of view since a student who is inactive during the first week test is likely to obtain poorer learning results and can therefore be prompted on the effects of his or her behaviour immediately.

Table 4

Learning Results of Active and Inactive Students

	Inactive	Active	Total
n	263	263	526
grade	6.297	7.347	6.822
	(0.100)	(0.089)	(0.070)
stdev	1.623	1.436	1.619
success	0.722	0.886	0.785
	Standard errors are between parentheses		
Δ grade	(8.800; 0.000)		
Δ success	(4.842; 0.000)		
eq. variance	(1.277; 0.024)		

About the previous analysis, one could argue that the active students are the qualitatively better students and that they would score higher anyhow. In order to examine whether this is the case, we use the result of the diagnostic test as a measure for the entry-level knowledge of the students. 490 of the sample of 526 students made the diagnostic test. We divided the group of students in two equally sized groups based on their score for the diagnostic test (below and above the median score). We assume that students in the group with below median diagnostic test score (we refer to this group as below) have a lower entry level of knowledge than the students in the group with the above median diagnostic test score (above). Table 5 shows the results for the active and inactive students within both the above and below groups.

Table 5

Learning Results of the Active and Inactive Students Corrected for Entry Level Knowledge

	Below		Above	
	Inactive	Active	Inactive	Active
n	162	83	121	124
grade	6.044	6.710	7.018	7.750
	(0.122)	(0.138)	(0.149)	(0.120)
stdev	1.552	1.261	1.638	1.336
success	0.691	0.807	0.826	0.935
	Standard errors are between parentheses			
Δ grade	(3.610; 0.000)		(3.828; 0.000)	
Δ success	(2.052; 0.020)		(2.661; 0.004)	
eq. variance	(1.515; 0.019)		(1.503; 0.013)	

Let's focus on the results for the below median group in table 5. Within this group of students, the active ones obtained an average grade of 6.710 whereas the inactive ones obtained a lower average grade of 6.044. The difference is significant with a p-value of 0.000 as the Δ grade row indicates. The standard deviation of the final grade is again lower for the active students than for the inactive students (1.261 versus 1.552, being a significant difference with a p-value of 0.020). The active students in this group were more likely to pass the exam than the inactive students (80.7% versus 69.1% being significant with 98% confidence). Apparently, the active students in the group of students who scored less than median in the diagnostic tests obtained significantly better learning results. More interestingly, this result is also apparent in the group of students who scored above median in the diagnostic tests; all differences are significant.

From these results we conclude that the impact of the online tests is robust with respect to the entry-level knowledge of the students. Being active in the online tests would increase the success rate with 12% for the group of students with below median score for the diagnostic test and with 11% for the group of students with above median score for the diagnostic test. Being active yielded on average a 0.666 higher grade for the below median students and a 0.732 higher grade for the above median students.

We basically assume that students with a below median diagnostic test score have less prior knowledge compared to students with an above median test score. Expected and desired behavior of those students with a below median test score is active participation in the weekly tests. As can be seen in table 5 out of the 245 students who scored below median on the diagnostic test, 162 are inactive with respect to an expected number of 141.5 inactive students. On the other hand, students who scored below median on the diagnostic test score are mostly active students (observed number of active students is 124, compared to 103.5 expected active students). The association is tested[7] and highly significant ($\chi2$=14.061 with p=0.000). This result suggests that students with an above median test score and assumed to have a higher course entry level are more likely to participate actively in the weekly tests, compared to students who have a below test score and therefore could benefit more of active participation in digital tests. A lesson that could be learned here is that the communication to students with a below median test score should be changed, the students should be made aware of the importance and impact of the weekly tests.

To further examine the robustness of our result, we analyzed the results using a different indicator for student entry-level knowledge. The students examined in this paper are Bachelor 1 students and just graduated from secondary school. On these schools, two types of mathematics courses are taught: mathematics A and B. Mathematics A consists primarily of statistics whereas mathematics B is more analytical. Out of both courses, mathematics B is most in line with the course we examine here. So we assume that mathematics A students (we refer to them with MA) are likely to have less prior knowledge than the mathematics B students (MB). From 434 students in the sample of 526 students, we know the mathematics course they followed at the secondary school. Table 6 presents the results.

Table 6

Exam Results of the Students Corrected for Entry-Level Knowledge

	MA		MB	
	Inactive	Active	Inactive	Active
n	155	87	102	90
grade	5.999	6.626	7.338	7.936
	(0.120)	(0.142)	(0.139)	(0.107)
stdev	1.492	1.326	1.409	1.012
success	0.690	0.816	0.902	0.967
	Standard errors are between parentheses			
Δ grade	(3.372; 0.000)		(3.405; 0.000)	
Δ success	(2.261; 0.012)		(1.860; 0.031)	
eq. variance	(1.266; 0.114)		(1.938; 0.001)	

Table 6 shows that more students have done mathematics A than mathematics B (242 instead of 192). Being active increased the average score for the exam with 0.627 and the success rate increased with 12.6% for the mathematics A students and the increase in the average score was 0.598 with 6.5% higher success rate for the mathematics B students. Again, we conclude that all students who participated actively in the online tests obtained a higher score for their exam and were more likely to pass, where mathematics A students did benefit more from activation than mathematics B students.

The content of mathematics A corresponds less with the course content compared to mathematics B, therefore we assume mathematics A students to have the lowest course entry level of both groups. Looking to the numbers of inactive and active students in table 6 a similar situation as was described in case of the diagnostics test scores occurs. Active participation in the weekly tests of mathematics A students is expected, but relatively many mathematics A students are inactive (155 inactive students while 143.3 inactive students expected), and on the other hand relatively many mathematics B students are active (90 active students compared to 78.3 expected active students). Again, we can conclude that the communication should change, the mathematics A students should know the effect of active participation in the weekly tests. We analyzed the impact of the online tests while correcting for entry-level knowledge in two ways: by the score for the diagnostic test and the type of mathematics course the students followed at the secondary school. It could be that some cross effects influence the results. To check for these potential cross effect of entry-level knowledge, we examined the results where we distinguish both in the secondary school mathematics course and the score for the diagnostic test. Table 7 shows the results.

Table 1

Exam Results of the Students Corrected for Entry-Level Knowledge

	Below				Above			
	MA		MB		MA		MB	
	Inactive	Active	Inactive	Active	Inactive	Active	Inactive	Active
n	95	50	39	19	49	33	54	67
grade	5.852	6.438	6.838	7.416	6.320	7.033	7.744	8.118
	(0.152)	(0.167)	(0.135)	(0.217)	(0.225)	(0.249)	(0.128)	(0.119)
stdev	1.478	1.182	1.453	0.841	1.574	1.430	1.271	1.407
success	0.674	0.800	0.846	0.947	0.735	0.879	0.944	0.970
	Standard errors are between parentheses							
Δ grade	(2.596; 0.005)		(1.912; 0.031)		(2.125; 0.018)		(1.534; 0.064)	
Δ success	(1.697; 0.045)		(1.306; 0.096)		(1.697; 0.045)		(0.692; 0.245)	
eq. variance	(1.564; 0.043)		(2.985; 0.008)		(1.212; 0.286)		(0.816; 0.777)	

Table 7 shows that the differences are less significant than before due to the smaller sample sizes. To discuss the results, let's assume that the Above / MB group of students have the most prior knowledge. For this group, the average grade and the success rate increased but surprisingly also the standard deviation of the exam grades. All of these differences are not significant, however. Apparently, being active had no significant effect on those students with the assumed highest level of prior knowledge. In all the other groups, we find significant results at the 95% confidence level (although the increase in success rate is significant only with 90.4% for the Below / MB group and the difference in variance is not at all significant for the Above / MA group). Apparently, activation has the largest impact on those students with the lowest assumed prior level of knowledge. The weekly online tests clearly helped to bridge gaps in the prior knowledge of the students. We therefore conclude that adding weekly online personal tests to a course with weekly group work significantly increases learning results, especially for those students who has less prior knowledge.

After doing the above described analysis, we conclude that activation by weekly tests leads to better students performance on the final course exam. One could argue that the active students are probably just the more motivated students, and therefore more involved in the course in general. We have insight in motivation of the students as they were asked to participate in a survey at the start of the academic year. The survey is developed by Arnold and Straten (2012) and aims to gather information about students' study choices and motivation. One of the questions in the survey was to give a grade (between 1 and 10) to indicate how motivated the student is to do the bachelor study 'Economics and Business Economics' (1 means lowest motivation and 10 highest motivation). We use the answers on this question as a measure for self-assessed motivation. Table 8 shows the self-assessments for the inactive and active students.

Table 8

Motivation Grades

	Inactive	Active	Total
n	92	77	169
grade	8.370	8.208	8.296
	(0.108)	(0.148)	(0.089)
stdev	1.035	1.301	1.163
	Standard errors are between parentheses		
Δ grade	(0.883; 0.190)		
eq. variance	(1.580; 0.018)		

Table 8 shows that in total 169 students out of the sample of 526 answered the question; 92 inactive students and 77 active students. The inactive students have a somewhat higher motivation grade compared to the active students. The inactive students gave themselves on average 8.370 on a 10-points scale and the active students 8.208, this difference is not significant as the Δ grade row indicates. The standard deviation of the motivation grades for the inactive students is 1.035 and for the active students 1.301. Row eq. variance shows that this difference is significant, so the standard deviation for the inactive students is significantly lower than for the active students. Based on these data we conclude that the active students are not just the more motivated students.

Conclusion

In this study we examined the effect of individual weekly tests in addition to group work on student learning. We made a distinction between students who participated actively in all weekly tests and students who did not based on their grades obtained for the tests. We find that active students obtain significantly higher exam grades and are significantly more likely to pass the final exam. To test the robustness of this result, we controlled for students' course entry level based on their secondary school mathematics level and results for a diagnostic test made on the first day of the course. Controlling for entry-level knowledge, we found higher exam grades and higher pass rates for active students compared to inactive students. Furthermore, while comparing the motivation grades the students gave themselves at the beginning of the year, we found no difference in motivation between the inactive and active students. We conclude that offering weekly tests in addition to weekly group work has a significant positive effect on student learning.

Notes.

1. A 5.5 is typically seen as a sufficient grade.
2. Recall that the students could redo the test after receiving feedback on the first test. It is therefore likely that a student who makes both attempts passes the test with a sufficient grade of 5.5 or higher.
3. A two-sample unpooled t-test with unequal variances.
4. A two-proportion unpooled z-test.
5. A two-sample F test for equality of variances.
6. See the discussion following Table 2 for the interpretation of the test results.
7. Chi-squared test for goodness of fit.

References

Angus, S.D., & Watson, J. (2009). Does regular online testing enhance student learning in the numerical sciences? Robust evidence from a large dataset. *British Journal of Educational Technology*, *40*(2), 255-272.

Arnold, I.J.M., & Straten, J.T. (2012). Motivation and math skills as determinants of first-year performance in economics. *The Journal of Economic Education*, *43*(1), 33-47, 2012.

Brothen, T., & Wambach, C. (2001). Effective student use of computerized quizzes. *Teaching of Psychology*, *28(*4), 292-294.

Charman, D., & Elmes, A. (1998). A Computer-based Formative Assessment Strategy for a Basic Statistics Module in Geography. *Journal of Geography in Higher Education*, *22*(3), 381-385.

Cluskey Jr., G.R., Hodges, C.W., & Smith, S. (2006). The impact of online quizzing on students success in an Introductory Financial Accounting class. *Journal of College Teaching and Learning*, *3*(7), 13-18.

Daniel, D.B., & Broida, J. (2004). Using web-based quizzing to improve exam performance: lessons learned. *Teaching of Psychology*, *31*(3), 207-208.

Griffin, F., & Gudlaugsdottir, S. (2006). *Using online randomised quizzes to boost student performance in Mathematics and Operations Research. Information Technology Based Higher Education and Training, ITHET '06. 7th International Conference on*, 481-489.

Haberyan, K.A. (2003). Do weekly quizzes improve student performance on general Biology Exams? *The American Biology Teacher*, *65*(2), 110-114.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* London: Routledge, 2009.

Kibble, J. (2007). Use of unsupervised online quizzes as formative assessment in a medical physiology course: effects of incentives on student participation and performance. *Advances in Physiology Education*, *31*(3), 253-260.

Steenhuis, H.J., Grinder, B., & de Bruijn, E.J. (2009). The use(lessness) of online quizzes for achieving student learning. *International Journal of Information and Operations Management Education*, *3*(2), 119-148.

Varsavsky, C. (2004). Can online weekly quizzes contribute to learning in mathematics? *Proceedings of the 9th Asian Technology Conference in Mathematics*. Singapore: ATCM.