Shailendra, P., Prageet A., Parul G., Diptiranjan M., Ratri P., Rebecca R., & Sumita Sindhi (2018). Online Education: Worldwide Status, Challenges, Trends, and Implications. *Journal of Global Information Technology Management*, 21(4), 233–241, doi: 10.1080/1097198X.2018.1542262

Shea, P. (2007). Bridges and barriers to teaching online college courses: A study of experienced online faculty at 36 colleges. *Journal of Asynchronous Learning Networks*, 11(2), 73–128.

Shenoy, V., Mahendra, S., & Vijay, N. (2020). COVID 19 lockdown technology adaption, teaching, learning, students engagement and faculty experience. *Mukt Shabd Journal*, 9(4), 698–702.

Simon, Edwige (2012). *The impact of online teaching on higher education faculty's professional identity and the role of technology: The coming of age of the virtual teacher*. University of Colorado at Boulder: ProQuest Dissertations Publishing.

Sumrall, J. (2002). *Factors which infuence faculty attitudes and perceptions of distance education in analytical subject areas*. Dissertation Abstracts International, 65(06), 2081A.

Thatcher, J. B., Loughry, M. L., Lim, J., & McKnight, D. H. (2007). Internet anxiety: an empirical study of the efects of personality, beliefs, and social support. *Information and Management*, 44(4), 353–363.

Thomas, S. M., and Cathy R. (2020). Education, The Science Of Learning, And The Covid-19 Crisis. *Prospects*, 49, 87–90. doi: 10.1007/s11125-020-09468-z

Toktarova, V & Semenova, D. (2020). Digital pedagogy: analysis, requirements and experience of implementation. *Journal of Physics: Conference Series*, 1691, 1-5. doi: 10.1088/1742-6596/1691/1/012112

West, R., Waddoups, G., & Graham, C. (2007). Understanding the experiences of instructors as they adopt a course management system. *Education Technology Research & Development*, 55(1), 1-26.

Wingo, N. P., Ivankova, N.V., & Moss, J.A. (2017). Faculty perceptions about teaching online: Exploring the literature using the technology acceptance model as an organizing framework. *Online Learning*, 21(1), 15-35. doi: 10.24059/olj.v21i1.761

Xhaferi, G., Farizi, A., & Bahiti, R. (2018). Teacher'attitudes towards e-learning in higher education in Macedonia Case study: University of Tetovo. *European Journal of Electrical Engineering and Computer Science*, 2(5), 14-17. doi: 10.24018/ejece.2018.2.5.26

Yusuf, Bibi N., and Jihad A. (2020). Are we Prepared Enough? A Case Study of Challenges in Online Learning in A Private Higher Learning Institution During The Covid-19 Outbreaks." *Advances in Social Sciences Research Journal*, 7(5), 205–212. doi: 10.14738/assrj.75.8211

PESOVSKI Ivica
Brainster Next College, North Macedonia

KLASHNINOVSKI Aleksandar
Brainster Next College, North Macedonia

MAKESKA Anja Marija
Brainster Next College, North Macedonia

THE EFFECT OF FREQUENT TESTING ON STUDENT PERFORMANCE

Abstract: This study aims to evaluate the effect frequent testing has on students following computer programming courses in an online environment. The experiment consisted of 26 students divided into two groups: 12 students who took tests after each lesson during the module introducing them to computer programming and 14 students who only took the final exam. The first group of students took 14 tests over the course of ten weeks. All 26 students took the final test at the end of the module. The test presented after each lesson to the first group of students consisted of 10 multiple choice questions related to the material covered in the class. The final exam consisted of 30 multiple choice questions, 10 completely new questions, 10 of the least correctly answered and 10 of the most correctly answered questions during the frequent testing phase. Students received immediate feedback on the test score and could see the correct answer to each question after submitting the quiz. The results show that students who took tests after each class

scored 13% higher on the final exam than those who only took the final exam. This result holds true when comparing the final exam scores on new questions only, on questions already seen and all together. The evident difference between the two study groups is statistically significant since the difference is large enough to secure higher grades for one of the groups.

Introduction

Evaluating student achievement during schooling is vital to the student experience. Students' destinies can be tailored by their grades during their schooling life, whether it is a primary school, a high school, faculty, or beyond. We often witness scenarios where students from different schools with the same GPA scores drastically differ on various entrance exams and student competitions. The current educational system is set up so that students get their knowledge tested periodically, most often at semestral exams which are held twice a year. This means that students must study throughout the whole semester or course duration and then get their knowledge evaluated on a single test conducted in a single day. Even the brightest students can have a bad day and fail to show their knowledge on this particular day which then acts as a huge disappointing factor and can pave the students learning path in a completely different direction than desired. On the other side, mediocre and even bad students can get lucky on that single occasion and get a good result, which can later lead to a lot of frustration for the students, the teachers and educational institutions if these students happen to choose a career path for which they are not motivated enough or even not capable of acquiring.

Although the problem is clear and people in charge of providing educational programs are well aware of it, no particular solution is put to practice in most schools. And the solution is pretty straightforward – instead of assessing students once per semester or per course, introduce a system of frequently testing their knowledge and forming their grades as a result of their consecutive learning and effort put into their studies. Experiments regarding which exam frequency works best for various courses have been done, and different periods have been proposed throughout the past decades [1,2,3]. Yet, even if many experiments showed that frequent testing improves student achievement and that even students report them as being helpful, rarely frequent testing is put into practice. The most probable reasons for this are the logistical barriers. First, the educational institution willing to consider frequent testing will need to prepare appropriate tests for each class, which shouldn't be an issue since many other materials are required for each class to be successfully conducted. But, the professor should then, after each class, check and grade each student's test, provide feedback and deliver the results back to students. This can

mean that a new job position may be required for specifically doing this work, knowing how much effort conducting even semestral exams takes, including delivering the tests, monitoring the students while the test lasts, reviewing the results and giving individual feedback to each student. Educational technology is by definition "the combined use of computer hardware, software, and educational theory and practice to facilitate learning." [4]. It understands incorporating technology in the education process to enhance the performance of either of its stakeholders, from students to teachers, parents, administration, external collaborators, etc. It doesn't mean that specific software or hardware needs to be used for a process to be considered as part of the educational technology. Even technology's theoretical or heuristic implementation can be assumed to be part of the ed-tech world. Despite this, most of the time an institution or process is qualified as ed-tech if it incorporates some kind of software or hardware solutions during the studies [5]. One of the main goals of educational technology is to improve students' learning performance by incorporating technology into the educational process [6]. It can do so by providing different tools students can use during their studies. All kinds of learning management systems (LMSs)

already exist which optimize students' experience with the complete learning process. One of the features that can easily be implemented in most of the LMSs publicly available or in the custom-built ones is the option to test students. Especially questions whose answer is easily validated by computers, like multiple choice or yes-no questions or questions expecting a specific number, word or sentence as an answer. Open-ended questions are more complex to be checked by computer and they require either the professors to check them or some kind of peer-to-peer checking and scoring mechanism. New machine learning (ML) techniques using NLP (Natural Language Processing) have already proven to be able to check open-ended questions [7, 8] and score them even better than humans without any kind of bias, a possibility for corruption, etc. The very good thing about these models is that once developed, they can easily be adapted to serve for different courses and in different languages [9,10]. This means that a single trained ML model can grade students' answers in all courses taught during the study's duration, expertise incomprehensible for a single human being.

Methodology

The methodology used in this research is quite straightforward. A class of students was divided into 2 groups. The first group was tested after each class for 2 months and then took the final exam. Before taking the final exam, the participants of this group were tested 14 times. The second group consisted of students that only took the final exam. The students were aware of the experiment and they got to choose which group they will fall into. The 14 tests that were given to the first group were presented to students right after the class finished and after their colleagues from the second group left the class. The classes were conducted online, same as the testing. For this purpose, a digital platform for testing was built. Each test consisted of 10 questions derived from the material that was just taught during the class, so the students were supposed to keep attention during the class in order to score better on the test after the class.

The final exam was realized on the same platform as the individual tests, only it consisted of 30 questions. Of those, 10 questions were pulled from the questions that students from the first group answered mostly correctly during the frequent testing phase. Then, the next 10 questions in the final exam were selected from the questions that students from the first group most often answered incorrectly. Finally, the remaining 10 questions on the final exam were completely new and no student was already familiar with any of the questions in this group. All of the questions in each of the tests were multiple choice questions. This means that each question had suggested answers and students had to choose what they think is the correct answer. The students were aware that only one of the suggested answers was true. No negative points were awarded if someone answered incorrectly, which means that students could randomly select an answer for a question and still have a chance of answering it correctly. After the students answered all questions and submitted their solutions, they were presented with immediate feedback about which of the questions were correct and which were incorrect. Because they were still online with their instructor, they could ask for further explanations about certain questions and the instructor would explain them after all the students had submitted their tests. The test was not time-limited, but the students and the instructor stayed online after the class until all students finished their test and the maximum time for test completion was 15 minutes (derived from the log files).

The methodology for analyzing the results is comparing the scores students from both groups got on the final exam. In this research, we are not analyzing the scores on the individual tests. We are only interested in discovering whether frequent testing helped students score better on the final exam.

Results and Discussion

The results that each student got on the final exam and info on how many tests each has taken are shown in Table 1. In the table we can see that out of the 12 students who participated in the first group, 9 took all 14 tests, 2 of the students missed one test on the days when they weren't able to attend the class and one student only took part in 8 tests and then took the final exam.
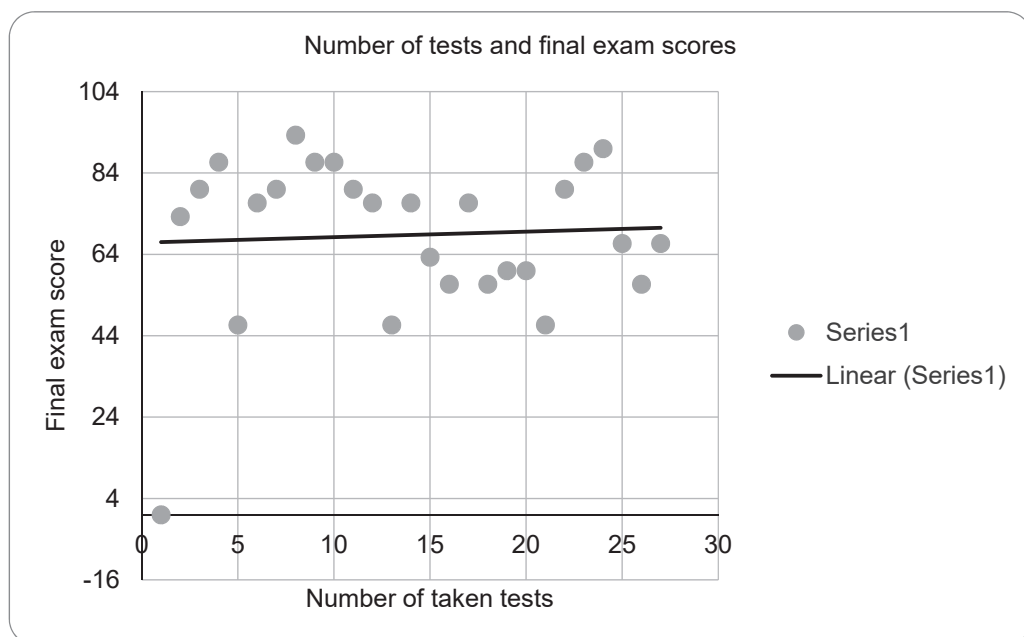
Table 1
Tests taken and final exam score

Student	Student group	# tests	Final score
Student 1	Frequent testing group	14 + final exam	73.3
Student 2	Frequent testing group	14 + final exam	80
Student 3	Frequent testing group	13 + final exa	86.67
Student 4	Frequent testing group	8 + final exa	46.67
Student 5	Frequent testing group	14 + final exam	76.67
Student 6	Frequent testing group	13 + final exa	80
Student 7	Frequent testing group	14 + final exam	93.33
Student 8	Frequent testing group	14 + final exam	86.67
Student 9	Frequent testing group	14 + final exam	86.67
Student 10	Frequent testing group	14 + final exam	80
Student 11	Frequent testing group	14 + final exam	76.67
Student 12	Frequent testing group	14 + final exam	46.67
Student 13	Final exam only group	final exam	76.67
Student 14	Final exam only group	final exam	63.33
Student 15	Final exam only group	final exam	56.67
Student 16	Final exam only group	final exam	76.67
Student 17	Final exam only group	final exam	56.67
Student 18	Final exam only group	final exam	60
Student 19	Final exam only group	final exam	60
Student 20	Final exam only group	final exam	46.67
Student 21	Final exam only group	final exam	80
Student 22	Final exam only group	final exam	86.67
Student 23	Final exam only group	final exam	90
Student 24	Final exam only group	final exam	66.67
Student 25	Final exam only group	final exam	56.67
Student 26	Final exam only group	final exam	66.67

A visual representation of the data shown in Table 1 can be seen in Figure 1.

Figure 1
Correlation between the number of tests and scores on the final exam

What we can see from the results in the table and from the graph is that there is a linear correlation between the number of tests taken and the scores on the final exam. The results on the final exam are definitely not deterministic by the number of tests one has taken. We can see that based on the second group of students who only took the final exam – there are students in this group who answered most of the questions correctly (Student 22, Student 23). And in both groups, the least successful student has the same score of 46.67 points. If the number of tests unconditionally influenced the success rate on final exams, we would have expected the minimum score for the students from the frequent testing group to be higher than the scores of the worst students from the second group, yet this did not happen.

Conclusion

In the results section, a detailed overview of each student's scores is given. There are 2 findings arising from this research that can be drawn out from the results based on the dataset that was used. The first conclusion is that the number of tests presented to a student doesn't necessarily mean that the student will do better in any case. We can conclude this based on the same lowest scores in each of the groups and the excellent scores students from the second group got on the final exam, despite not participating in any of the previous tests. So, frequent testing can improve student performance but does not guarantee it.

The second conclusion is that students being tested frequently score better than their colleagues who only took one test when we observe them as a group. The finding visible in the results section that students that were tested frequently during the course length scored 13% higher on the final exam is a significant difference and can contribute toward students getting a whole higher grade in an academic context. In combination with the previous conclusion, this conclusion can be formed to express that being in a group that is frequently tested increases one's chances of scoring better results on the final exam but does not guarantee that happening.

Future Work

The results presented in this paper are based on an experiment performed on only 26 students. This is a pretty small dataset and one of the ongoing processes while writing this paper is replicating the same experiment with more students. In the future, we will be able to do this analysis on a larger representative sample once more data is gathered. The findings of this study are expected to be confirmed, if not even amplified, as the sample size grows. Another analysis that can be done on the current dataset, but will be even more interesting as the sample grows, is to analyze the student's performance on final exams divided into question groups. As mentioned ins the methodology section, the final exams consist of 30 questions of which 10 are questions that were answered correctly by most students during the frequent testing phase, 10 are questions that were mostly answered incorrectly and the last 10 questions are completely new. An analysis of these question groups can show whether students perform better on questions that they have already seen, whether complex questions whose feedback was already given to students are easier the second time they come around and whether the frequent testing improves student achievement only on familiar questions or on new questions as well.

The final thing that arises as future work from this research is to deliver a survey to each student, especially to each student from the first group of students (who were tested regularly), asking them to evaluate their experience with the frequent testing, whether it was stressful for them, whether they liked it and whether they believe that the frequent tests contribute toward their education. Based on the results from this survey and analysis of the students' stress levels regarding having frequent tests can be evaluated and the process for conducting tests can be adapted and optimized regarding the findings in the survey.

References

Bangert-Drowns, R. L., Kulik, J. A., & Kulik, C. L. C. (1991). Effects of frequent classroom testing. *The journal of educational research*, *85*(2), 89-99.

Stiggins, R. (2005). From formative assessment to assessment for learning: A path to success in standards-based schools. *Phi Delta Kappan*, *87*(4), 324-328.

Rezaei, A. R. (2015). Frequent collaborative quiz taking and conceptual learning. *Active Learning in Higher Education*, *16*(3), 187-196.

Educational technology. (2022). In *Wikipedia*.

https://en.wikipedia.org/w/index.php?title=Educational_technology&oldid=1098451312

Weller, M. (2018). Twenty years of Edtech. *Educause Review Online*, *53*(4), 34-48.

Donahoe, B., Rickard, D., Holden, H., Blackwell, K., & Caukin, N. (2019). Using EdTech to enhance learning. *International Journal of the Whole Child*, *4*(2), 57-63.

Krithika, R., & Narayanan, J. (2015, August). Learning to grade short answers using machine learning techniques. In *Proceedings of the Third International Symposium on Women in Computing and Informatics* (pp. 262-271).

Singh, G., Srikant, S., & Aggarwal, V. (2016, August). Question independent grading using machine learning: The case of computer program grading. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 263-272).

Çınar, A., Ince, E., Gezer, M., & Yılmaz, Ö. (2020). Machine learning algorithm for grading open-ended physics questions in Turkish. *Education and information technologies*, *25*(5), 3821-3844.

Ariely, M., Nazaretsky, T., & Alexandron, G. (2022). Machine learning and Hebrew NLP for automated assessment of open-ended questions in biology. *International journal of artificial intelligence in education*, 1-34.