MÉTODOS DE AVALIAÇÃO NO ENSINO A DISTÂNCIA DE CÁLCULO: UM RELATO DE EXPERIÊNCIA

Methods of Assessment in Remote Teaching of Calculus: a Report of Experience

Xu Yang [yang@ic.ufal.br]
Instituto de Computação, Universidade Federal de Alagoas, Maceió, AL, Brasil
Xiao-Chuan Liu [lxc1984@gmail.com]
Instituto de Matemática, Universidade Federal de Alagoas, Maceió, AL, Brasil

Recebido em: 21/10/2022
Aceito em: 09/05/2023

Resumo

O método de avaliação é um dos tópicos mais importantes na educação matemática. Durante a pandemia do COVID-19, todas as atividades de ensino no Brasil foram remotas. Na Universidade Federal de Alagoas, como centenas de outras universidades semelhantes, nossa disciplina de cálculo 2 para alunos formados em Engenharia de Computação também é 100% online nessas circunstâncias especiais. Queremos aplicar um método de avaliação que possa representar razoavelmente bem o desempenho dos alunos e reduzir a probabilidade de trapaça, que se tornou um problema sério com atividades remotas. Este trabalho é composto pela avaliação quantitativa da avaliação formativa e somativa dos alunos. Quatro grupos de estudantes de Engenharia de Computação participaram de todas as avaliações. Especialmente na avaliação somativa, as questões são distintas para cada aluno. Os dados incluem 40 listas e 8 exames. Fizemos análises estatísticas sobre os dados. A partir dos dados e da análise, concluímos que durante o ensino remoto ainda podemos avaliar os alunos de forma justa e honesta. Isso exige que o professor pense mais em como elaborar o exame com base na realidade e na natureza da matéria. Quanto à matemática, existem muitas maneiras de testar a capacidade de computação e compreensão dos alunos.

Palavras-chave: método de avaliação; matemática; ensino remoto; avaliação formativa; avaliação somativa.

Abstract

The assessment method is one of the most important topics in mathematical education. Due to the COVID-19 pandemic, all teaching activities in Brazil are being conducted remotely. At the Federal University of Alagoas, like hundreds of other similar universities, our course, Calculus 2, for Computer Engineering majors was also conducted entirely online due to these special circumstances. We want to apply a method of assessment which can fairly well represent the performance of the students and reduce the likelihood of cheating, which became a more serious problem with remote activities. This work is composed of quantitative evaluation of the formative and summative assessment of students. Four groups of students majoring in computer engineering participated in all the assessments. Especially in the summative assessment, the questions are distinct to every student. The data includes 40 lists and 8 exams. We did some statistical analysis on the data. Based on the data and analysis, we have concluded that during remote teaching, it is still possible to assess students fairly and honestly if exam questions are designed properly. This requires the teacher to think more on how to design the exam based on the reality and the nature of the subject. As for mathematics, there are many ways to test the ability of computation and comprehension of students.

Keywords: assessment method; mathematics; remote teaching; formative assessment; summative assessment.

INTRODUCTION

In March 2020, due to COVID-19, it was announced that all the schools including primary schools, high schools and universities have to move to emergency remote teaching and assessment. This significant change posed a lot of challenges for both teachers and students. In July 2020, we did a survey among students on the situation of access to computers and the internet. There were 407 students who replied to the question sheet. As the results indicated, 5 (1.2%) students did not have a computer, 14 (3.4%) students occasionally used computers, 56 (13.8%) students shared a computer with the family and 332 (81.6%) students had their own computers. Regarding internet access, 6 students (1.5%) could only access the internet via their cell phones, 199 students (48.9%) could only access it through residential internet, and 200 students (49.1%) could access it through both cell phone and residential internet. About the quality of the internet, there were 2 (0.5%) students whose internet was very bad so that they have difficulty with reading the email and opening a website, 60 (14.7%) students whose internet was not good, so that they can only open a website, but cannot watch a video or participate an online meeting, 345 (84.8%) students have good internet so that they can participate online meeting and watch videos. Moreover, besides the problem of computers and the internet, there were also other potential problems which are not listed in the survey, such as whether the students have cameras or microphones. What is more, the electricity system in Maceio is also very bad. Many times, after a heavy rain, there is no electricity for several hours. Non-expected interruption of electricity is also very common due to unknown reasons. All these problems can create many challenges in teaching and assessment.

In our case, our course was calculus towards engineering students. Most of the time, mathematical courses can be given in a very simple way. The classes mainly consist of lectures with the help of boards and chalks. In particular, the course does not have practical parts, such as experiments, neither need any equipment or machines from the university while teaching and learning. So during the pandemic, the teaching part can proceed as before. Now the only change is that we give the course on-line, through computer, camera and maybe a digital writing pad. But what we were more concerned about was the assessment of the students' work. Assessment links the teachers and students during the process of the study and also can be an important motivation for the students to keep studying. Assessment should be done in a very formal way. On one hand, the assessment should reflect the true status of the students' learning process. But now this relates to the honesty of the students. We thought about asking the students to use cameras during the exams. But it turned out to not be practical. Not all the students have a device such as a camera or even cellphone so that we can full time check them when they are doing an exam. The other aspect of the assessment is about the major of the students. Our students are from computer engineering, who concentrate more on computation and application of mathematics. Calculus is one of the basic courses in mathematics for them. There are now many websites and software that can basically do most calculations such as do integrals in a perfect way. It is very difficult to check whether the students indeed solved the problems by themselves, or whether they did them with the help of some high-techs online. With all these concerns, we developed some kind of evaluation method. With the benefit of the nature of the subject, our method works very well, and does not give too much extra work for the teachers. At least in Ufal, there are other courses on mathematics for the engineering students, including probability and statistics, linear algebra, analytic geometry, and complex analysis. The current forms of the assessment for these other courses are a combination of quiz, written exam, weekly assignments and so on. These courses clearly share similarities with our calculus courses, and therefore our method might be helpful also.

Universidade Federal de Alagoas (UFAL), as the best university in the state of Alagoas, is very serious with the evaluation of every course. There is a general rule of evaluation published in the website, which can be found here: https://ufal.br/estudante/graduacao/normas/avaliacao_do_rendimento_escolar. In this website, it is announced that the evaluation of school performance will be done through the following procedures.

- Bimonthly Assessment (AB), 4 (four) times per academic year;
- Final Exam (PF), when applicable;
- Completion of course work.
- Each Bimonthly Assessment (AB) should be limited to the contents developed in the respective bimester and will result from more than one assessment instrument.
- In each two-month period, the student will have his/her grade in the respective Bimonthly Assessment (AB) calculated considering the total points obtained divided by the number of evaluations programmed and carried out by the discipline.
- In each subject, the student who achieves a grade lower than 7 (seven) in one of the 4 (four) Bimonthly Assessments (AB), will be entitled, at the end of the academic year, to be reassessed in the one in which he obtained the lowest score, prevailing, in this case, the revaluation grade.
- During the pandemic, this general rule of assessment was not changed.

Here we have several points to stress. Every semester, we should give the students two grades AB1 and AB2 in the system. If the sum of AB1 and AB2 is between 10 and 14 (inclusive), indicating that at least one of AB1 and AB2 is less than 7, the student will have to complete an additional evaluation to replace the lower score. If after the replacement, the sum is still less than 14, but more than 10, then the student has to participate in a final exam. In this report, we do not talk about the final exam part, nor the reevaluation.

THEORETICAL BACKGROUND

Methods of assessment are important because they can reveal the status of the study and help students and teachers to improve the overall learning outcomes. Generally, there are two models of assessment, formative assessment and summative assessment (Kerka & Wonacott, 2000). Formative assessment normally happens along with the learning process (Harlen & James, 1997), which "is essentially feedback, both to the teacher and to the pupil about present understanding and skill development to determine the way forward" (Harlen & James, 1997). Therefore normally it only takes a small portion of the final grade. In contrast, summative assessment is to evaluate the students at the end of the semester. It should involve the whole range of content. Therefore, it is highly recommended that the hybrid of the two evaluation methods be carried out (Buchholtz, Krosanke, Orschulik, & Vorhölter, 2018). As for mathematics, especially before COVID-19, closed-book assessment is used more often, because it is easier for the teachers to design the questions and correct the answers. However, there are already many discussions about this form of evaluation, because this requires the students to remember too many formulas and theories which sometimes are not necessary and also at the same time not efficient (Ní Fhloinn & Carr, 2017). Online assessment, on the other hand, developed rapidly during these years. Croft, Danson, Dawson and Ward (2001) reported their experience of using computer-assisted assessment (CAA) in engineering mathematics. During COVID-19, alternative methods other than closed-book assessment were applied. Fitzmaurice and Ní Fhloinn (2021) investigated the assessment methods by doing an online survey with 257 teaching staff in 29 countries. They reported that 80.5% give online assessments, and 38.5% give formative and summative assessments. The format of formative assessment is homework or assignments, multi-choice questions or online tests, written assessment and upload, timed assessments, oral assessments, etc. The format of summative assessment is an open-book exam (untimed), assignments, online assessments, open-book exam (timed), multi-choices questions, projects, oral exam, etc. Olawale, Hendricks, and Mncube (2021, November) did a study on students from South Africa. They suggested

that during the crisis, multimodal technologies are important in assessment. Online assessment should "focus on problem-solving, decision-making abilities, as well as the cognitive demands."

There are many works on the models of assessment in mathematical education. However, not all of them are suitable for engineering students. Ní Fhloinn and Carr (2017) proposed several formative assessment ideas for engineering students, including an online approach. Davis, Harrison, Palipana, and Ward (2005) described the HELM (Helping Engineers Learn Mathematics) learning resources and assessment regime. Sazhin (1998) described a way of doing formative assessment, which is called spot tests. That is, in a semester, the teacher randomly selects 3 classes without informing the students. At the end of each class, the students are asked to solve some questions within 20 minutes given by the teacher. Spot tests can be a motivation for students to study consistently. What we did with a similar idea is that in every class, we raise some questions related to what we learned in the former classes. If they answer correctly, we give them extra grades in AB1 or AB2.

METHODOLOGY

Regarding assessment, we adopted both formative and summative assessments. For formative assessment, we choose to give one assignment weekly. More precisely, we post a list on google classroom and the students have one week to finish it. The problems in the lists consist of mostly standard exercises, with possibly one last question a bit more difficult. About the summative assessment, we have one mid-term exam and one final exam. Every exam has two parts. One part is an open-book exam, which is limited to 3 hours and is relatively simple. The other part is an assignment, which is more difficult and we allow the students to collaborate and discuss. In this report, we focus on the exam part. The data we used to analyse are the grades of the every-week assignment and the two exams[1]. We don't disclose the names and other personal information of the students. As a basic mathematics course for engineering students, our bottom line of teaching is that the students can understand the basic concepts and do the calculations. The exams' questions focus on

- understanding the basic definitions of integrals, sequences and series
- being able to calculate integrals with techniques such as the substitution of variables, integration by parts and so on.
- Being able to test the convergence or divergence of series using the integral test, comparison test, alternating series test, ratio test, and root test.
- Being able to apply the integrals and series to solve practical problems such as calculating the areas and volumes.

Regarding summative assessment, it is important to minimize the possibility of cheating. For this reason, the questions for the students are all different. At the time of the exam, we send the questions to every student by email and they also submit the answers by email personally.

The student part:

- One day before each exam, email the teacher and confirm participation in the exam.
- At the time of the exam, after receiving the exam, answer the questions, take photos and send them back to the teacher.

The teacher part:

[1] There was no Informed Consent Form. The research project does not have an ethics committee. Moreover, the researchers always preserve the anonymity of the students. The information used in this work is only the grade. This work proceeds and explicitly assumes and exempts Acta Scientiae from any consequences arising therefrom, including full assistance and possible compensation for any damage resulting from any of the research participants, in accordance with Resolution No. 510, of April 7, 2016, of the National Health Council of Brazil.

- Before the exam, design the question sheets. While receiving the email from the students about the confirmation, generate more question sheets by slightly changing the questions. Program the email of the question sheet to every student who will participate in the exam.
- At the time of the exam, confirm with the students if they have received the exam.
- After the exam, correct the answers and send the results back by email.

If the students have doubts about their grades, they can also reply to the email. In this way, all the activities can be traced in the emails.

RESULTS AND ANALYSIS

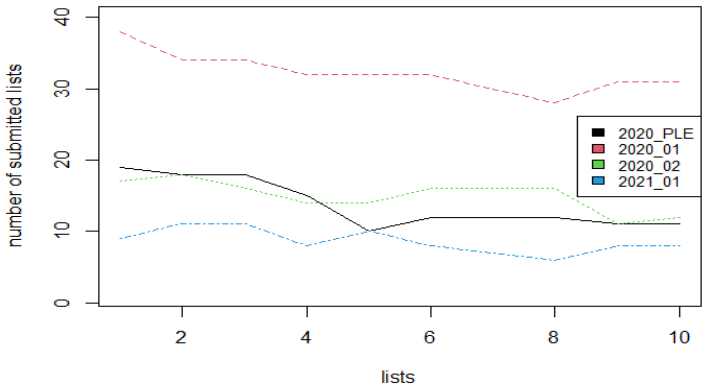
In this section, we will report the statistical analysis of the method of assessment. During the COVID-19 lockdown, 100% online courses on calculus 2 are given in four semesters, which are 2020-PLE (September, 2020 – January, 2021), 2020-01 (February, 2021 – June, 2021), 2020-02 (July, 2021 – October, 2021) and 2021-01 (October, 2021 – March, 2022). The contents of this course are given in the following (also see Stewart (2015)):

Integrals, Areas and Distances, the Definite Integral, the Fundamental Theorem of Calculus, Applications of Integration, Areas between Curves, Volumes, Volumes by Cylindrical shells, Techniques of Integration, Integration by Parts, Trigonometric Integrals, Trigonometric Substitution, Integration of Rational Functions, Strategy for Integration, Improper Integrals, Polar Coordinates, Infinite Sequences and Series, Sequences, Series, the Integral Test and Estimation, the Comparison Tests, Alternating Series, Absolute Convergence and Tests, Ratio test, Root test, Strategy for Testing Series, Power Series, Representations of Functions, Tayor and Maclaurin Series.

Every semester has 15-17 weeks and every week we have two classes with each class lasting 100 minutes. The total number of students in each class are 35 (2020-PLE), 59(2020-01), 31(2020-01) and 12(2021-01). But every semester, several students inscribed themselves but never showed up in class, nor participated in any evaluation. We estimate this proportion with the number of the most submitted list divided by the total number inscribed. The proportions for every semester are 54%, 61%, 45% and 91%. We can see that in the semester 2021-01, our class is a relatively small group, but most of the students actively participated in the class and submitted the assignments. Next, we show in Figure 1 the situation of formative assessment, which happens every week in google classroom and all the students have the same question sheet. We can see from Figure 1, that every semester, the number of submitted lists declines. At the beginning of every semester, the students had more time and were more motivated to finish the assignments. After a few weeks, some students may give up and skip the assignments. By the end of the semester, some might realize the importance of doing and submitting exercises, and therefore start to submit again. This may also be related to the content of the course. That is, it depends on whether or not the students like certain parts of the course.

Figure 1

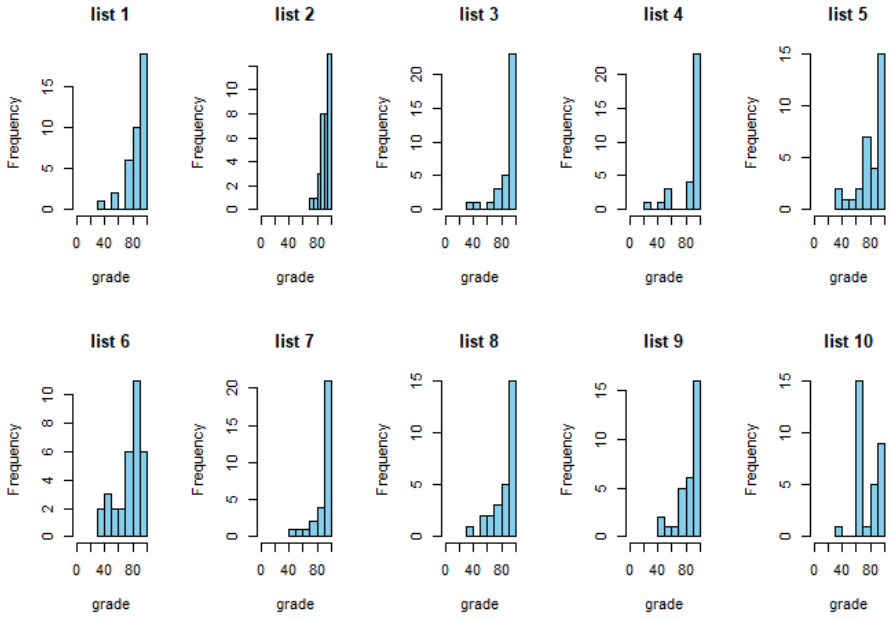
The Number of submitted assignments in every semester.



In Figure 2, we show the histogram of grades of 10 lists in semester 2020.01, which has the most students.

Figure 2

Histograms of the grades in every list.



From Figure 2, we can see that except for lists 6 and 10, the grades concentrate on 90-100. Most of the students get grades above 90 in the formative assessment. Even with lists 6 and 10, there are still many students who get grades of more than 80. The same result is confirmed with the other datasets in semesters 2020-PLE, 2020-02 and 2021-01. We will show the frequency data in the appendix.

Now we move on to summative assessment. Our summative assessments include 2 exams, a mid-term exam and a final exam. The summative assessments are followed by the method described in Methodology. To sum up, the questions in the summative assessments are very similar to the lists, however, the form is by sending an email instead of being announced by google classroom. In this way, we can make sure that the students get distinct but similar questions. We ask the students to finish their own questions and submit them within 3 hours. Note that, their final grade, AB1 and AB2 are composed of 30% (or 50%) of formative assessment and 70% (or 50%) of summative assessment.

The following are some examples of exams (see Figure 3, Figure 4, Figure 5 and Figure 6). We can see that in every figure, we show two questions or question sets, which are highly similar but are different. In Figure 3, there are two parts. The left and right parts have 4 questions for calculating integrals. The two (a) questions are to test the integral of the absolute value of x. But they have different upper limits and the same lower limits. In these two questions, different upper limits do not complicate the calculation of the integrals in any way. With this method, we can also create more similar integrals with different lower limits or with different lower and upper limits. Of course, if we analyze the two integrals further, we can argue that the first integral is a little bit easier than the second one because the function in the first integral is an odd function and the interval is symmetric with respect to zero, so the integral is zero. As for the second integral, we still need to discuss and calculate it. The two (b) questions are to test the integrals of trigonometric functions and the technique of substitution of variables with different lower and upper limits. The two (c) questions are designed to test the integration of exponential functions and the technique of variable substitution, while the two (d) questions aim to test the substitution of trigonometric functions or variables. In Figure 4, the questions are both to calculate the area using an integral. In both questions, the two functions are highly similar so that the same person can solve the two questions using the same time. In Figure 5, we present two examples of questions about substituting variables. Although these two questions are also about the substitution of variables, in comparison with questions (b) in Figure 3, the questions in Figure 5 can test better if the students really understand the technique and can use it correctly. In practice, we noticed that several students succeeded in solving questions (b) in Figure 3, but were stuck on questions in Figure 5. Figure 6 is an example of similar questions about series. It is much easier to construct similar questions about series to guarantee that they are different and with the same level of difficulty. Note that the six questions covered all the tests we studied, such as the integral test, comparison test, alternating series test, ratio test, and root test. In every example, every pair of corresponding questions have the same aim and topic, but with slightly different coefficients. In this way, there is no necessity for the students to consult with each other, not to say they have limited time to finish their own exams. In all four semesters, we didn't receive any complaints from the students about the difficulty and complexity of their exams.

Figure 3

The first example of calculating integrals.

$(a)\int_{-1}^{1}(2x|x|)dx$ $(b)\int_{1}^{2}sen(\pi t/2)dt$ $(a)\int_{-1}^{3}(2x-3|x|)dx$ $(b)\int_{0}^{1}sen(\pi t/3)dt$

$(c)\int_{0}^{1}\sqrt{1+2x}dx$ $(d)\int\frac{x}{\sqrt{1-4x^2}}dx$ $(c)\int_{0}^{3}\sqrt{1+6x}dx$ $(d)\int\frac{x}{\sqrt{1-2x^2}}dx$

Figure 4

The second example of calculating areas.

(2) Esboce a região delimitada pelas curvas indicadas. Decida quando integrar em relação a x ou y. Calcule a área da região

$$y = x^2 - 3x - 1, y = x + 4$$

(2) Esboce a região delimitada pelas curvas indicadas. Decida quando integrar em relação a x ou y. Calcule a área da região

$$y = x^2 - 2x - 2, y = x + 2$$

Figure 5

The third example of questions in the exam.

(2) Se f for contínua e $\int_0^2 f(x)dx = 6$, calcule $\int_0^{\pi/2} f(2cos\theta)sen\theta d\theta$.

(2) Se f for contínua e $\int_0^2 f(x)dx = 6$, calcule $\int_{\pi/2}^{\pi} f(2sen\theta)\cos\theta d\theta$.

Figure 6

The fourth example of questions in the exam.

(3)Teste a convergência ou divergência das séries

$$(a)\sum n^{100}(0.99)^{2n} \qquad (b)\sum \frac{(-1)^n}{n+1}$$

$$(c)\sum \frac{\cos 2n}{n^3+1} \qquad (d)\sum \frac{(n!)^3}{(3n)!}$$

$$(e)\sum ne^{-2n} \qquad (f)\sum \frac{3^n}{4^{n+1}-n}$$

(3)Teste a convergência ou divergência das séries

$$(a)\sum n^{100}(0.99)^{n} \qquad (b)\sum \frac{(-1)^n}{2n+1}$$

$$(c)\sum \frac{\cos n}{n^3+1} \qquad (d)\sum \frac{(n!)^2}{(2n)!}$$

$$(e)\sum ne^{-n} \qquad (f)\sum \frac{2^n}{3^{n+1}-n}$$

In Table 1, we list the number of students who participated in the exam. We can see that in the semester 2020-PLE, 26 students (74% out of total number of students) participated in the first exam, however, only 11 students (26% out of total number of students) participated in the second exam. This is the first online semester, the students may still need time to adapt. Besides, the number of students who passed the exam with grades over 7 also dropped significantly with only 1 student passing the exam. But on the other side, in semester 2021-01, which is the last semester with courses online and also has the least students, we have the most proportion of students who participated in the exam and also passed the exam. But we also note that, in every semester, both the number of students who participated and passed in the exam declined.

Table 1

The number of students who participated in the first and second exam.

	2020-PLE	2020-01	2020-02	2021-01
Exam 1	26 (74%)	47(75%)	18(45%)	9(75%)
Grade > 7	15 (58%)	22 (47%)	12 (67%)	6 (67%)
Exam 2	11 (26%)	39(62%)	18(45%)	8(67%)
Grade > 7	1 (9%)	9 (23%)	6 (33%)	3 (38%)

The following figures are the histograms of grades. Note that, the questions of the exams are also similar to the exams before the COVID 19-pandemic. At that time, we only gave the students 2 hours to finish the questions and hand in the answers. During the pandemic, we give the students 3 hours with the extra time to take photos or to digitize their answers and then use the computer and send it. We consider that maybe there are students who do not have a computer or internet at home and they may have to go somewhere else where it might not be very comfortable and convenient to do the exam. So some extra time is necessary for them. Ultimately, our goal is to ensure that the exam remains an effective tool for informing both students and teachers about the state of learning, rather

than serving as a form of competition. In fact, if the students submit the answer only after 3 hours, we still accept and correct them, only with a small discount for the grades.

In Figure 7, we show the histograms of two exams in 2020-PLE, 2020-01, 2020-02 and 2021-01. For more details on the frequency, see the appendix. To contrast with Figure 7, in Figure 8, we plot the histograms of the two first lists every semester.

Figure 7

The histogram of two exams in 2020.PLE, 2020.01, 2020.02 and 2021.01

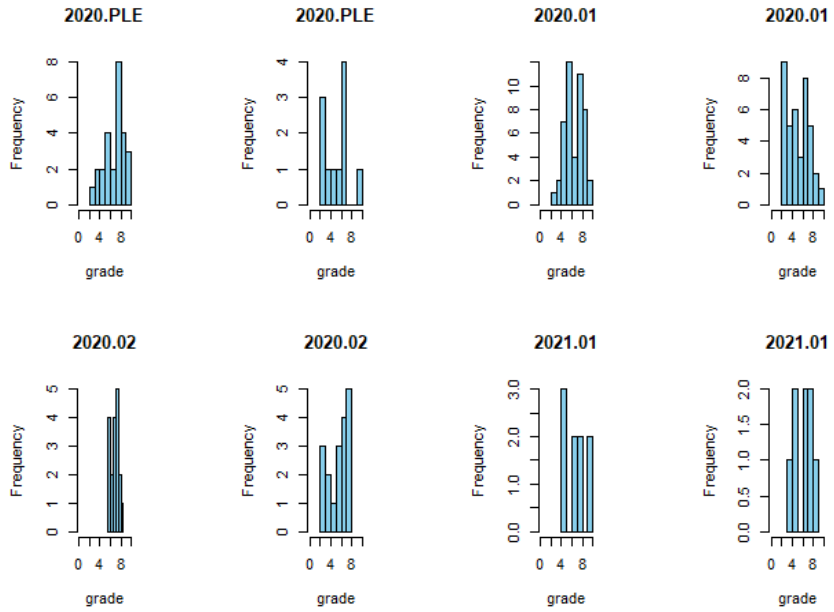
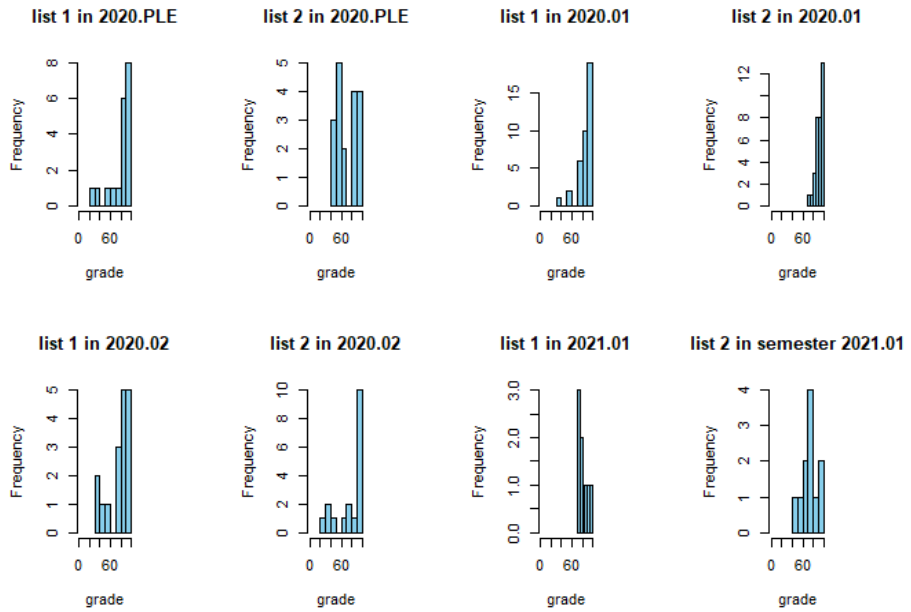


Figure 8

The histogram of two lists in semester 2020.PLE, 2020.01, 2020.02 and 2021.01.



From Figure 7 and Figure 8, we can see that the students generally get lower scores on the exams. One reason is that they may have a restricted amount of time to complete the exam. It is possible that even when they do the exam at home, they may still feel nervous and anxious. Every semester, there may be one or two students who forgot to attach a page of their answers to the email. However, the timing is not the most important factor, because most of the students send their emails before the deadline. There are even some students who send the email half an hour before the deadline. So we would like to consider another important reason which is that the questions of the exam problems are unique for each of them. They cannot discuss or even copy other fellow colleagues' answers. Of course, here we do not need to forbid the students to consult online. Some questions, such as calculating the integrals can be solved on many websites perfectly. About this part, we will discuss further in the Conclusion section. Based on the difference between Figure 7 and Figure 8, we believe that our summative assessment is relatively fair and reliable.

If we study the data further, we can find more interesting questions to think about. For every semester, we calculate the correlation of the grades of all ten lists between every two students. We want to check the students who have a correlation of more than 0.9. We find 4 students with correlations of more than 0.9, we list the difference in their grades of lists and exams below in Table 2.

Table 2

Two pairs of students have a high correlation with their lists. Note that since the lists have a total grade of 100, we scale to 10 and the exams have a total grade 10.

	Student 16 - Student 49	Student 41 - Student 48
List 1	0	-0.5
List 2	0	0
List 3	0	-0.2
List 4	-0.3	-0.4
List 5	1.3	-0.4
List 6	0.7	0.2
List 7	0.1	-0.5
List 8	2	-0.3
List 9	1	0
List 10	0	0
Exam 1	2	0.2

Exam 2	3	-4.1
correlation	0.92	0.96

From Table 2, we can see that the grades of lists of student No.16 and student No.49 have correlation 0.92 and student No.41 and student No.48 have correlation 0.96. But there is a big difference in their second exam. The differences between Student No.16 and Student No.49 are 2 and 3. The scores of student No.41 and student No.48 share a similar situation. The difference in exam 2 between Student No.41 and Student No.48 is 4.1. Note that their grades of list 6 to list 10 are very similar. Here we use the command cor() in R to calculate the correlations. It only computes the correlation between columns with no missing values. If there are any missing values, like "na", it returns "na" as well. Now we consider the situations of missing values. The command cor() in r has two options for dealing with missing datas, one is "use=complete.obs" and the other one is "use=pairwise.complete.obs". The "use=complete.obs" is not applicable, because there are many students who only submitted one or two lists. We choose to add the command "use=pairwise.complete.obs". We find that 27 pairs of students have correlations of more than 0.9. In Table 3, we show an example of the difference of their grades. From Table 3, we can see that Student No.48 had relatively good grades both in formative assessment and summative assessment. Student No.9 also didn't participate in the first exam (from Table 2, we know that Student No. 48 participated in both exams, but Student No. 9 didn't participate in the first exam, so the difference here is NA) and performed poorly in the second exam. In this case, the grades of lists 6 to 10 are not trustable. The same happens for Students No.24 and No.27. Student No. 24 participated in both exams but Student No. 27 didn't participate in the second exam. We also calculated the correlation coefficients for other semesters. In Table 4, we list the results.

Table 3

The difference of grades of lists and exams of students 9 and 48, and students 24 and 27.

	Student 24-Student 27	Student 9-Student 48
List 1	NA	NA
List 2	-3	NA
List 3	NA	0
List 4	NA	NA
List 5	1	NA
List 6	-0.5	1
List 7	0	0
List 8	NA	0.5

List 9	1	0
List 10	0.5	0.3
Exam 1	1.2	NA
Exam 2	NA	-4.8
correlation	0.92	0.94

Table 4

The number of pairs of students with correlation coefficients of the grades more than 0.9 in semester 2020-PLE, 2020-01, 2020-02 and 2021-01.

	2020-PLE (pairs of students)	2020-01 (pairs of students)	2020-02 (pairs of students)	2021-01 (pairs of students)
Ignore NA value	0	4	18	0
Consider NA value	2	27	52	1

In Table 4, Semester 2020-02 has the most number of pairs of students who have high correlation coefficients. Here we also want to emphasize that the high correlation coefficients do not suggest that the two students copy from each other. Some more complicated reasons might be expected. As we see in Table 2, with such results, it is probably a good idea to contact students No.49 and No.41 to discuss the problems in the exam and try to understand more about their personal reasons.

CONCLUSIONS

In this work, we report our assessment methods for the course calculus 2 during the COVID-19 pandemic. We choose to do a combination of formative and summative assessments. The formative assessment encourages the students to keep pace with the subject. The summative assessment evaluates the students at the end of the semester. Regarding the summative assessment, we developed our own method of evaluation, which is to give the students distinct questions with similar difficulty levels. Therefore, during the exam, they can focus on their own questions. Because of the nature of this subject, mathematics, it is feasible to generate 20-30 similar versions of exams. This assessment method can reduce the likelihood of cheating and is very fair for the students. Although it creates additional workload for teachers, such as preparing questions and grading answers, the increased effort is manageable if the number of students is not too large. In our experience, the extra effort is completely worthwhile. After four semesters teaching, we have the following concerns:

- If it is possible, we should make the class as a small group with 10-15 students. In this way, we can notice the study situation of every student more clearly. For example, in semester 2021-01, we only had 12 students, but 9 of the students actively participated in the assignments and exams. The similarities of the lists are also very low, that is, the correlation of the grades of lists between the students is very low. The teacher can also get to know the students during the class and the correction of assignments. For a small group

of students, we can make the class and the assessment in many different forms. For example, we can add some interviews with every student as a part of the assessment, which can be more trustable and reliable. More specifically, about every assignment, instead of only giving them grades, we can arrange some extra classes to explain the problems and with emphasis on the mistakes they have made. After the exam, we can also correct the exam with every student online together. During the correction, we can ask the student to explain more about their ideas.

- Another advantage of our summative assessment method is that it is also easier to do reevaluation. If some students have doubts about their grades and want to participate in the reevaluation, we can send them another version of the exam that differs from what they have received before, but again, similar. Along the way, we build our library of exams for future use, which includes 20-30 different question sheets with similar difficulty levels.
- The reliability of an online assessment is questioned, especially when the questions to the students are the same. As we can see in the Figures and Tables, the grades of formative assessment and summative assessment are not totally consistent. In all four semesters, the grades of formative assessments are relatively high and concentrated in the interval of 90-100 points. However, the results of the summative assessment concentrated on relatively lower grades. For the formative assessment, we actually encouraged students to discuss and form a study group to finish every assignment. In every class, we also asked the students if they had any questions about the assignments. As expected, it was found that students who performed better tended to participate more in class and ask more questions, while those who performed poorly did not participate in the class at all. For the summative assessment, on the other hand, we are very serious about the honesty of the students. That is the reason why we design this system of assessment. But this doesn't totally solve the problem. For example, we ask questions about calculating integrals as in Figure 3. There are several websites like https://www.integral-calculator.com/, https://www.wolframalpha.com/calculators/integral-calculator/ and https://pt.symbolab.com/solver/integral-calculator which can solve integrals with all the details. While correcting the answers, we already see some clues that certain students copy from these websites. In this situation, we consider some modifications of the questions, such as restricting solving the integral using a certain method. For example, in Figure 3 question (d), the websites solve it using substitution of variables, so we can ask the students to solve the question using substitution of trigonometric functions. But this may receive some complaints from the students, like, they may claim that they are good at substituting variables and there is no reason to forbid a certain method. Therefore, a better solution for this would be to give more questions like in Figures 4 and 5, instead of those in Figure 3. This is also a better way, in that, questions in Figures 4 and 5 combine some applications of calculating integrals. Maybe the students can still use the websites to solve the questions, but at least they have to first understand the questions and give a correct expression of the integral. We saw several students who only answered the questions in Figure 3 but did not answer anything in the questions in Figures 4 and 5. Moreover, as our students are majoring in computer engineering, questions in Figures 4 and 5 are in fact much better because they are designed to be sort of applications and not only pure computations. On the other hand, designing proper questions like in Figures 4 and 5 also requires much more time.
- The regularization of evaluation of our university gives several opportunities for the students to participate in the assessment, such as AB1, AB2, reevaluation and final exam, which are good for the students to have better grades, but there are students who take advantage of this to not participate in the class and not do any assignment. In our course, AB1 and AB2 comprise 30% of formative assessments in every weekly list and 70% of (two) exams. But the grade of reevaluation alone can replace the grade of AB1 or AB2.

So some students took advantage of this according to this regulation. Their strategy was the following. They did not participate in the classes, nor did any of the assignments given weekly. They only participated in the two exams and got grades in AB1 and AB2, and then they tried their best to participate in the reevaluation to achieve a better grade. In this way, they "saved" a lot of time from doing the assignment and still had a chance to pass. Due to a lack of control over honesty in the exam, it is very much easier for them to cheat in one or two exams. Under this circumstance, we were considering some changes to the reevaluation part. If possible, we suggest the reevaluation be in the form of an interview or a presentation, so that the teacher can know better about the situation of the students. However, this idea still had the disadvantage that we could not do the interview if we had too many students who needed to do the reevaluation. Interviews and presentations are time-consuming and do not have clear criteria to give the grade. Therefore, if we teach in a small group, then an exam combined with interviews and presentations should be more applicable.

- The COVID-19 pandemic brought us many challenges and thoughts in teaching and assessing. After the pandemic, we are now back to mostly presential teaching with at most 40% online activities. Currently, in our courses, if we want to add some extra classes, like exercise classes, we can choose to teach online and record as well. This is somewhat new because, before the pandemic, we never did online teaching. The teachers did not even have a device to do online teaching, like a digital pad. However, we still prefer the summative assessment to be presential. If there is any student who cannot participate in the exam in person, we can alternatively send her/him a similar version of the exam, as we did during the pandemic. Compared to the closed-book exam before, we also want to bring more technology into the assessment format. Some new ways of teaching and assessments can be kept after the pandemic.

AUTHORS' CONTRIBUTIONS STATEMENTS

X-C.L. and X.Y. conceived the presented idea and developed the theory. X.Y. adapted the methodology to this context, created the models, performed the activities, collected the data and analyzed the data. All authors actively participated in the discussion of the results, reviewed and approved the final version of the work.

DATA AVAILABILITY STATEMENT

The data supporting the results of this study will be provided by the corresponding author, X.Y, upon reasonable request.

REFERENCES

Buchholtz, N. F., Krosanke, N., Orschulik, A. B., & Vorhölter, K. (2018). Combining and integrating formative and summative assessment in mathematics teacher education. *ZDM*, *50*(4), 715-728.

Croft, A. C., Danson, M., Dawson, B. R., & Ward, J. P. (2001). Experiences of using computer assisted assessment in engineering mathematics. *Computers & Education*, *37*(1), 53-66.

Davis, L. E., Harrison, M. C., Palipana, A. S., & Ward, J. P. (2005). Assessment-driven learning of mathematics for engineering students. *International journal of electrical engineering education*, *42*(1), 63-72.

Fitzmaurice, O., & Ní Fhloinn, E. (2021). Alternative mathematics assessment during university closures due to Covid-19. *Irish Educational Studies*, *40*(2), 187-195.

Harlen, W. & James, M. (1997). Assessment and learning: differences and relationships between formative and summative assessment. *Assessment in education: Principles, policy & practice*, *4*(3), 365-379.

Kerka, S. & Wonacott, M. E.. (2000). *Assessing Learners Online. Practitioner File*.

Ní Fhloinn, E. & Carr M. (2017). Formative Assessment in Mathematics for Engineering Students. *European Journal of Engineering Education 42* (4): 458–470.

Olawale, B. E., Hendricks, W., & Mncube, V. (2021, November). Rethinking Assessment in Mathematics Teacher Education Programme during COVID-19 Pandemic. In *Proceedings of ICERI2021 Conference* (Vol. 8, p. 9th).

Sazhin, S. S. (1998). Teaching mathematics to engineering students. *International Journal of Engineering Education*, *14*(2), 145-152.

Stewart, J. (2015). *Calculus*. Cengage Learning.

APPENDIX

In semester 2020-PLE, the frequency of grades of every list and exam.

	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
list 1	1	1	0	1	1	1	6	8
list 2	0	0	3	5	2	0	4	4
list 3	0	1	1	3	3	4	5	1
list 4	1	0	0	2	3	2	5	2
list 5	1	0	1	0	4	2	0	2
list 6	0	0	1	1	0	3	3	4
list 7	0	0	1	0	0	2	2	7
list 8	0	0	0	0	3	4	4	1
list 9	0	1	1	1	1	1	1	5
list 10	0	0	1	1	0	2	1	6
exam 1	1	2	2	4	2	8	4	3
exam 2	3	1	1	1	4	0	0	1

In semester 2020-01, the frequency of grades of every list and exam.

	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
list 1	0	1	0	2	0	6	10	19
list 2	0	0	0	0	0	2	11	21
list 3	0	1	1	0	1	3	5	23
list 4	1	0	1	3	0	0	4	23
list 5	0	2	1	1	2	7	4	15
list 6	0	2	3	2	2	6	11	6
list 7	0	0	1	1	1	2	4	21
list 8	0	1	0	2	2	3	5	15
list 9	0	0	2	1	1	5	6	16

list 10	0	1	0	0	15	1	5	9
exam 1	1	2	7	12	4	11	8	2
exam 2	9	5	6	3	8	5	2	1

In semester 2020-02, the frequency of grades of every list and exam.

	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
list 1	0	2	1	1	0	3	5	5
list 2	1	2	1	0	1	2	1	10
list 3	1	0	0	2	1	7	2	3
list 4	0	0	0	2	1	2	4	5
list 5	1	0	0	0	2	2	4	5
list 8	0	0	0	0	1	3	5	6
list 9	0	1	0	1	0	2	1	6
list 10	0	0	0	0	0	7	4	1
exam 1	0	0	0	4	6	8	1	0
exam 2	3	2	1	3	4	5	0	0
	0-50	50-100	100-150	150-200				
Lists 6 and 7	1	1	2	12	0	0	0	0

In semester 2021-01, the frequency of grades of every list and exam.

	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
list 1	0	0	0	0	0	5	2	2
list 2	0	0	1	1	2	4	1	2
list 3	0	3	1	0	1	1	0	5
list 4	0	2	0	0	5	1	0	0
list 5	1	1	0	0	1	1	2	4
list 6	0	1	0	0	1	1	3	2
list 7	0	0	0	0	0	1	3	3
list 8	0	0	0	0	1	0	1	4

list 9	0	0	1	1	1	1	0	4
list 10	0	0	0	0	0	3	3	2
exam 1	0	0	3	0	2	2	0	2
exam 2	0	1	2	0	2	2	0	0