Annals of Surgical Education

ISSN: 2692-7926

Research Article

Towards a "Liquid" Evaluation during the COVID-19 Pandemic: Challenges, Opportunities and Changes Arising from Reflections of Medical Students and Surgical Lecturers

Héctor Guadalajara Labajo[1,2], Barbara Soria-Juan[1,2], Saúl Fernández-González[1], Álvaro Palazón[2], Isabel Pascual Migueláñez[3], Olatz Lopez-Fernandez[1] and Damián García-Olmo[1,2]*

[1]Department of General Surgery and Digestive System Unit, Hospital Universitario Fundación Jiménez Díaz, Spain

[2]Department of Surgery, Universidad Autónoma de Madrid, Spain

[3]Department of General Surgery, La Paz University Hospital, Spain

Abstract

Background: The pandemic has challenged the traditional educational model of medical students, and it has introduced a new concept: 'liquid education'. This study aims to test a solution agreed on a webinar between students and lecturers to adapt the evaluation under a contract during the first lockdown.

Materials and methods: This action-research used a concurrent transformative mixed-methods design through an online debate, including 120 undergraduates and 9 lecturers to extract a consensus document to examine confined students. A follow up was done to test the exam pattern responses.

Results: The agreement was based on four principles: evaluation needed an environment of fairness, the minimum grade was protected, practical training required reformulation, and a continuous liquid evaluation strategy was essential in a future academic course. However, students showed scepticism to online teaching and evaluation and proctoring methods. A few exam response patterns were identical, detecting online cheating on exams.

Conclusion: To only evaluate online through a consensus is bound to fail. The exams within Universities are crucial and to connect academic and professional values. We provide reflections, recommendations, and examples about how to implement continuous liquid evaluation.

Keywords: COVID-19; Medical education; Surgical training; Liquid education; Online evaluation; Academic integrity; Online cheating on exams

Introduction

Higher education is considered a high-level form of adult education. Student participation and co-responsibility are critical factors in their training and successful academic and professional outcomes. Evaluation is one of the most relevant parts of the educational process, considered the starting and final point of the spiral of the learning-teaching method, and it is usually based on suitability, autonomy, and objectivity [1]. Indeed, students' perceptions of evaluation significantly influence their approaches to learning and studying. Although criticism exists around the traditional evaluation methods, critics tend to be sceptical of new systems [2]. Similarly, lecturers have concerns about the conventional system (e.g., standardized tests) and have reservations about the alternative evaluation approaches [2,3].

In 2000, Zygmunt Bauman created a visionary sociological construct that has acquired great recognition during pandemic times: 'liquid modernity [4], which uses the instantaneous metaphor malleability of a liquid against a solid world applied to contemporary societies. He postulates current societies are more fluid due to the dynamic nature, which produces continuous changes that require rapid transformations and difficult solid structures. The pandemic has shown evidence supporting this theory and how fluid our system can be, although the risks that Bauman postulated were present [5] (e.g., 'de-institutionalized' education due to lockdowns, lack of enough technical skills and devices, or capacity to react). The new environment required urgent adjustments towards a flexible and provisional liquid education, adapting methods and technologies to support evaluations while preparing the educational agents (i.e., students and lecturers) to change.

The first wave of the coronavirus disease 2019 (COVID-19) compromised academic engagement, especially in South-Europe. Internet technologies were not enough for the unexpected formative assessments and summative evaluation requirements during pandemic

Citation: Labajo HG, Soria-Juan B, Fernández-González S, Palazón A, Migueláñez IP, Lopez-Fernandez O, et al. Towards a "Liquid" Evaluation during the COVID-19 Pandemic: Challenges, Opportunities and Changes Arising from Reflections of Medical Students and Surgical Lecturers. Ann Surg Edu. 2021;2(1):1014.

Publisher Name: Medtext Publications LLC

Manuscript compiled: May 26th, 2021

***Corresponding author:** Héctor Guadalajara Labajo, General Surgery and Digestive System Unit, Hospital Universitario Fundación Jiménez Díaz, Madrid, 28040, Spain, E-mail: h.guadalajara@quironsalud.es; hector.guadalajara@uam.es

conditions, which showed to be underdeveloped [6], especially in medical education what contrasts with the inclusion of technologies in clinical practice [7], and the development of digital health (e.g., telehealth, eHealth). The human aspect of the educational processes (e.g., adaptive capacity in teaching and learning, academic integrity in evaluations) generated concerns. Before the outbreak, lecturers and students had already begun to discuss methods to adapt evaluations to contemporary realities as part of a mutual learning process. Both educational agents were aware of modifying procedures to improve learning measurement with validity using new methods adjusted to current technologies.

However, the pandemic exposed the deficiencies of the traditional evaluation model and the current use of technologies. For instance, it was not possible to use the physical university teaching and evaluation methods (e.g., paper-and-pencil exams with traditional Multiple-Choice Questions (MCQs) and Short Essay Questions (SEQs) due to the first strict lockdown. Furthermore, a Learning Management System (LMS) did not facilitate a quick fix of medical practical training within hospitals that were dealing with COVID-19 and coping with a lack of health and safety measures. Spain was one of the most severely hit countries in Southern Europe [8], in which the pandemic highly disrupted higher education. During this first lockdown, the authors managed to seize the opportunity presented by this new scenario as an impetus to overhaul the system in a way that the academic community had been calling for: to start applying liquid education through a rapid reaction and democratic process to adapt the evaluation. On 24 April 2020, a group of medical students and lecturers at the 'Universidad Autónoma de Madrid' (UAM) invited the university community to participate in the webinar "Evaluation during the COVID-19 Pandemic".

The present study aims to narrate the experience of this action-research inquiry to establish a set of consensus-based conclusions between both parties and to use the consensus result from the debate as a guide with an evaluation contract about examining medical students in pandemic times to prepare a new adapted way to perform evaluation differently through a continuous liquid evaluation in medical education.

Materials and Methods

Mixed methods design

The mixed-methods action-research design used was a concurrent transformative strategy with an expansion methodological aim, in which the qualitative method (i.e., firstly, a content analysis of the online debate, and secondly a few informal interviews with students who had identical patterns of responses in the exams) was seeking to explore different facets of the phenomenon of starting to transform evaluation through the other quantitative method (i.e., online survey to students participating in the webinar through a chat and in a follow up analysing the exam results). It was pretended, firstly, to extract the main concerns from both parts regarding this rapid transition and dynamic characteristics of the pandemic affecting the evaluation to achieve a consensus (through a qualitative phase within the webinar conversations that embedded a quantitative phase in the chat through an online survey). Secondly, to focus on behaviors during the online MCQ online evaluation to observe if the results were as usual academic years and protected academic integrity.

Participants and procedure

The debate was arranged by a group of nine lecturers and nine students in 'Introduction in Surgical Practice' at UAM with the support of the Spanish State Council of Medical Students (CEEM). The event was announced and promoted via LMS, social media platforms, and student forums. The debate was organized using the new platform "Operemos" (https://operemos.es/), which has social networking features, that facilitates the management of continuous training by healthcare professionals. By using the platform, lecturers can share real experiences, students can connect with accredited mentors, and patients can connect with experts. It facilitated the management of the event through an online registration process, but the webinar was broadcast via the Cisco WebEx meeting platform, a videoconferencing system used to hold online meetings with high digital video quality.

The event took place on 29 April 2020 with a maximum of 129 attendees (120 students and 9 lecturers). Additionally, the organisers held two open debates on the platform in which attendees participated *via* text chat and live questions. The debates were overseen by three moderators (i.e., researchers), who allotted speaking time equally among lecturers and students. Subsequently, a consensus document from students' and lecturers' perspectives (i.e., a guide in a non-face-to-face setting to online examine confined medical students) was generated through an agreement on 'Good Practice During Online Evaluations' with an evaluation contract before the final online exam.

Measures and analysis procedure

The debate content was analysed through a thematic content analysis of the transcription of the webinar, which is qualitative method frequently used in health research [9]. Responses to the 11-item survey were analysed using descriptive statistics. Firstly, this qualitative analysis used through an inductive extraction of codes grouped in subthemes, which were subsequently grouped in four themes representing the main findings categorised as the recommendations made by all panel lists (students and lecturers), which underwent a consensus-building process summarised and tagged by a support staff team (i.e., researchers). Secondly, in each of the response of all survey questions was computed its frequency and percentage. Thirdly, the data extracted from the exam of this course was statistically analysed using R (v 4.0.2), and the potential identical pattern of exam responses was analysed by the Copy Detect package (v 1.3) to compute response similarity indices for MCQ Tests (due to SAQs were checked by Turnitin). The analysis calculated different indices that showed the probability of the similarity of the pattern responses being random to estimate possible cheating on online exams. To do the calculations, each answer was considered correct, incorrect, or not answered. When an answer was incorrect, the option selected was not considered. The statistical analysis considered the probability of the coincidence of an identical exam submission in groups of two-to-two. In each pair of exams compared, the W index showed that similarity was random with a $p < 1.10\text{-}5$. For a .05 confidence level, every index returned the same result. With an n of 205 students examined, the probability of having two identical exams is 1 per 100,000 cases. Fourthly, as an additional evaluation was held, a set of informal short interviews were conducted with a few students regarding identical patterns of the first exam results.

Regarding ethics, the UAM lead panel lists provided information to students about the procedure of the webinar by means of various channels (e.g., brochure, email lists). The lecturers organized the webinar together with the students, the CEEM, and researchers. The survey was administered by the researchers before, during, and

at the end of the webinar through a function included within the chat in Webex. Each student responded to the questions voluntarily, anonymously, and confidentially in the webinar during around an hour and a half long. Anonymous feedback was provided after each response to the questions to let them know about the overall answers in each option to enhance the debate. Researchers asked the participants' consent for using the results to extract the common views of the evaluation through the information collected. Subsequently, the statistical analysis of the exam results was conducted by a university disciplinary committee, and the UAM research ethics committee supported the ethics procedure due to the action-research nature. According to the Good Clinical Practice principles, this study was carried out as defined in the International Conference on Harmonisation and full conformity with ethical standards. The Spanish and European regulations on the protection of personal data were also observed during the study.

Results

Themes and responses during the debate

Themes regarding to promote good practices in evaluations during COVID-19:

Results extracted from the lecturers and students who participated in the online debate through the chat are shown in Table 1.

In summary, regarding the first theme related to students and lecturers' psychological situation, academic evaluations were planned and administered in a high-stress environment with frequent anxiety, burnout, and above all, depression from both sides. The mental health problems discussed were mainly due to the pandemic, which disrupted participants' lives, education, professional responsibilities, thereby creating a need to modify the evaluation system rapidly. Other issues regarding the uncertainty surrounding their academic future and the dramatic situation that the country was experiencing, with relatives and acquaintances affected by COVID-19, led to the following first conclusion of the debate. In performing their duties, lecturers and students had to promote empathy (i.e., efforts to understand others' situation). Regarding evaluations, lecturers had to evaluate students in an environment of fairness rather than frustration.

Concerning the second theme about minimising the negative impact on students' academic record, a persistent consideration regarding the conundrum of academic evaluations during the pandemic was associated with the negative impact on medical students' academic record. It was discussed how it might have been mitigated if there had been an alternative model of continuous evaluation, e.g., a model to keep going academic processes in times of catastrophic situations such as the pandemic. However, the model did not exist in Spanish medical faculties beginning in 2020. Thus, the second conclusion was that the minimum mark received by a student during the pandemic should not be lower than the other marks on his/her transcript records. Hence, to not damage student academic record and future opportunities, which in medical students' case are essential due to the competitiveness of the profession and the relevance of this workforce in pandemic times.

The third theme was related to evaluating practical training, which is a key factor in medical education and surgery, and requires face-to-face assessments (e.g., in operating rooms). It is usually concentrated in the last years of medicine, and it is performed within teaching hospitals, which were highly disrupted by the first wave of this pandemic. The discussion covered the national experience versus international solutions during the first wave. The third conclusion was the practices performed in close European countries (e.g., the UK model) were near to Spanish reality on practical evaluation options, being environmental and cultural factors key in this issue. This aspect, however, was the most difficult one to include in the current evaluation, providing options to recover the practices after the academic course.

Lastly, regarding ethical issues and reflections in liquid evaluation, breaking the evaluation schemes emerged. Simultaneously, alternative options were proposed, such as creating new questions to conduct exams fairly and comprehensibly from both sides. Therefore, the last conclusion is the need to rapidly and fluidly adapt higher education using a liquid continuous evaluation, creating a fair and trustworthy environment of evaluation that will protect the human factor through consensus in how to conduct it, and academic integrity to avoid cheating on online exams. 3.1.2. Students' survey responses concerning evaluation during COVID-19.

Responses from the students participating in the chat during the webinar

Results extracted from the attendees who participated in the chat and answered the survey are shown in Table 2.

New online evaluation system *vs.* Traditional exam system

Ordinary exam method adapted to an online exam using Moodle after an evaluation contract:

The online exam for the course "Introduction to Surgical Practice" was held on 5 June 2020 using the UAM Moodle platform. Since it was a new experience for the university, the conclusions reached by medical lecturers and students during the debate were considered for the development of the exam.

The exam was divided into two equal and separate parts, each weighted as 50% of the final score. The first part consisted of two SEQs referring to a surgical procedure shown in a video recording. The second part consisted of forty MCQs with five answer choices, only one of which was correct. A correct answer was rewarded with one point of the total mark and a wrong answer penalized by subtracting 0.33 points from the total score. The exam's MCQs were sequential, i.e., the exam began with the first ten questions (of the total forty), and once the student had answered them, he/she would agree and move on to the next ten questions without the possibility of going back. The duration of the exam was 60 minutes. Students had previously signed the agreement on 'Good Practice during Online Evaluations' (i.e., evaluation contract) with a promise to sit the online exam.

Results of ordinary evaluations during COVID-19 pandemic

Of 205 students, 16 failed the exam, 53 students passed, 133 students had a remarkable grade, 3 students had an excellent mark, but no student obtained a distinction (Figure 1).

Patterns on the multiple-choice exam questions:

Exam questions are usually divided into three levels of difficulty: some questions are easy, while most of the questions are intermediate (i.e., if the student has studied, he/she should know the answers). 15% of the questions are more difficult and are often used to distinguish students who have the greatest control over the material. Surprisingly, when lecturers analysed the MCQ exam answers, they observed repeated response patterns in certain students (Table 3).

Table 1: Themes, subtheme/s, and quotations from students and lecturers who participated in the webinar "Evaluation in times of COVID-19".

Themes	Subthemes	Quotations from lecturers and students
1. Psychological situation	Online technologies	"The lecturer can also observe the stress experienced by a student when doing an exam online, as the limited timing, complexity of the situation, the insistence on new instructions to avoid cheating on online exams can increase students' distress' (P1: male lecturer)
		"Inequality, not all students have the same conditions, connection, etc." (Student from the chat)
	Pandemic per se	"Independently of the need of future changes, we need to evaluate now, in an environment of conflict and depression. Half of the students manifest this condition based on data, but lecturers' data too. In our course, half are sick -with COVID-19- and one has died. It is normal to be depressed in this situation. The lecturer and student need to accept the new environment we have with the pandemic. We need to evaluate fairly, not with frustration." (P8: male lecturer)
		"If mental health cannot be evaluated, it can cause discrimination -in evaluation- according to the circumstances -pandemic" (P3: female student)
	Tools for mental health solutions	"Maybe we need to translate the tools and good work done by psychiatrists and mental health staff working within hospitals in universities to medical students; online sessions are possible (…) tools can be provided to students to manage depression" (P6: female lecturer)
		"We need to limit the emotional load in the academic activities, to listen to all stakeholders who can help to understand how to manage this load and reduce it, as well as the level of anxiety" P2: male student)
2. Minimising the impact on students' academic record	Technological/ Management support	"It will be the same -evaluation- but from home, without a pilot test due to the exceptional situation. A ranking may be made to discriminate work positively. There are tools to evaluate and discriminate, such as medicalum (http://www.medicalum.com)" (P4: male lecturer)
		"Continuous evaluation should have been encouraged, even though we are late. A possibility of giving weight to all of the above (before the exam) (P4: male student)"
	Humanistic support	"Are we going to relax the standards? No, the standards should not be lowered, only adapted to the teaching performed -in pandemic times. Students have been empathetic with clinical lecturers, but they are not receiving the same teaching (P4: male student)
		"Evaluating, with fairness, but helping and being comprehensive (...) we need to avoid distress in students, universities will help them." (P7: male lecturer)
3. Evaluating practical training	Spanish medical practices to evaluate	"Even students voluntarily working in a hospital -during the pandemic- have not been able to receive classic practical training. How are students going to be evaluated in this context?" (Student from the chat)
		"There are other options, such as make up for this missed teaching outside the ordinary academic year with fewer students and smaller groups within hospitals, maybe in the summer period" (P7: female lecturer)
	International medical practices to evaluate	"There are two alternatives used in other countries: to provide practical knowledge through online activities which will enhance strategies such as clinical thinking; this is being used in countries such as Italy, the UK, and Austria; or to include the pandemic in the Medicine curriculum, and to teach medicine in COVID-19 times, as Denmark has done" (P5: female student)
		"In France exams are online but conducted within the University, students attend the exam physically managed using university computers. However, is the same to evaluate practical competences with to know how to practice medicine? In six years in Medicine, I have never had a practical exam. Before the pandemic times, this was a problem" (P6: female lecturer)
4. Ethical issues and reflections regarding liquid evaluation	Humanistic medicine and education	"Are we up to the task? Students live in the XXI century, we needed to adjust (…) the form of teaching and the evaluation system (…) we need to think about the challenge we live in and the future scenario. We need to interact more with our students and get to do a real continuous evaluation (…) to differently behave" (P1: male lecturer)
		"We are in an exceptional situation, consequently evaluation should also be exceptional (…) lecturers and students are in the same boat, it is fundamental to look for consensus and change the traditional concepts we have related to evaluation, and if we are wrong, we would be making mistakes together and learning from them." (Student from the chat)
	Liquid education	"Fluid education is in line in how to adapt all what we were trying to create in online teaching adjusted to students' different situations -personalised education. The University is their students" (P2: male student)
		"Medicine is a practical discipline; how do you evaluate that? How do you learn it? There is a disagreement between what we see in daily practice, what we teach and what we evaluate, so we must learn to better adapt to this situation through a "liquid" teaching" (P1: male lecturer)
	Liquid evaluation	"There is a need to adapt current evaluation to continuous evaluation by adjusting methods, workloads, and by reducing the weight assigned to the last exam" (P2: male student)
		"There is an obligation to make students feel that they are studying in a fair environment, so computer skills do not take away the maximum grade to whom it deserves it. In medicine, agents have been worried to get good grades -not to pass exams. Fair environment and do not help hackers" (P8: male lecturer).
	Academic integrity	"Students know technology well and are experts in breaking the rules (…) the exam can be known before being done" (P2: male student)
		"By now, we can only use traditional methods, and to adapt them to an online alternative. We must focus on establishing the rules for such an evaluation. It must be agreed between students and teachers, although the latter are the auditors. We are forgetting that the medical degree is a professional degree. We want good doctors!" (P8: male lecturer)

'P' is Panellist and 'X' is the number of the panellist, nine lecturers and nine students, plus students from the chat whose voices were provided by the researchers managing the chat.

a. Similarity identification:

In analysing student answers, three groups with different answering patterns were identified. Three possible results were assigned to analyse the patterns of responses: 0=wrong answer; 1=not answered; and 2=correct answer. These patterns were repeated by many other students in a high percentage of the exam without being identical.

b. Identifying identical exam submissions:

Seven exam submissions had an answer pattern that was identical with at least another submission. The distribution of identical exam submissions among the three different patterns was as follows: Pattern A with two identical submissions, Pattern B with two identical submissions, Pattern C with three identical submissions. Table 2 shows the answers of the seven identified students grouped by their answering pattern.

Table 2: Descriptive results of the survey from the webinar attendees.

Item	Responses	Fi (%)
1. Do you consider that the current situation of the pandemic represents an inflexion point in university education methods?	Yes, it is an opportunity to change the traditional model.	41/68(60)
	No, it is just a break in traditional education.	4/68(6)
	Unanswered.	23/68(34)
2. Do you think that all students currently have access to the internet and devices so that it is possible to carry out a completely online education?	Yes, every student has internet access and devices.	5/76(7)
	Most students have access to internet and devices.	40/76(53)
	Less than 50% of students have internet access and devices.	1/76 (1)
	A minority has access to the internet and devices.	0/76 (0)
	Unanswered.	30/76(39)
3. How do you consider your workload to be during the SARS-CoV-2 pandemic?	It has increased.	26/66(39)
	It has remained the same	7/66(11)
	It has decreased.	2/66(3)
	Unanswered.	31/66(47)
4. Do you think that the level of academic burden placed on students should be decreased during the pandemic?	Yes, I think so. The situation is exceptional.	20/81(25)
	The level of academic demand should not be decreased, but the teaching and evaluation methods should be adapted to this situation.	19/81(24)
	Unanswered.	42/81(52)
5. Do you consider that this course's average mark should not be lower than the student's average?	I agree.	25/65(38)
	I disagree.	3/65(5)
	Unanswered.	37/65(57)
6. Do you think online evaluation with video surveillance is appropriate?	Yes, I totally agree.	5/75(7)
	Only in specific cases.	24/75(32)
	I am afraid I must disagree due to privacy issues.	13/75(17)
	Unanswered.	33/75(44)
7. Do you think that continuous evaluation should be adopted at university?	Yes, I do.	42/75(66)
	No, I do not think so.	2/75(3)
	Unanswered.	31/75(41)
8. Do you think traditional exams can be adapted to current online evaluation platforms?	Yes, I do. All the exam formats can be done electronically.	29/73(40)
	No, I do not think so.	16/73(22)
	Unanswered.	28/73(38)
9. How do you value the academic year extension to recover the training practices lost during the pandemic?	It seems appropriate to me if the hospital situation allows it.	25/75(33)
	I think training practices can be recovered throughout the degree.	11/75(15)
	I think this year's training practices should be missed.	4/75(5)
	Unanswered.	35/75(47)
10. Do you think students' opinion, through their representatives, should be considered when making decisions regarding evaluation?	I do not think so. Lecturers can make these decisions.	0/79(0)
	Yes, although lecturers should make the final decision.	8/79(10)
	Yes, I do.	32/79(41)
	Unanswered.	39/79(49)
11. Do you consider lecturers as efficient as usual outside of face-to-face class?	Yes, they are just as efficient.	4/71(6)
	No, teaching is much more difficult.	21/71(30)
	No, but the level of teaching is acceptable.	15/71(21)
	Unanswered.	31/71(44)

'Fi' is frequency and '%' is percentage.

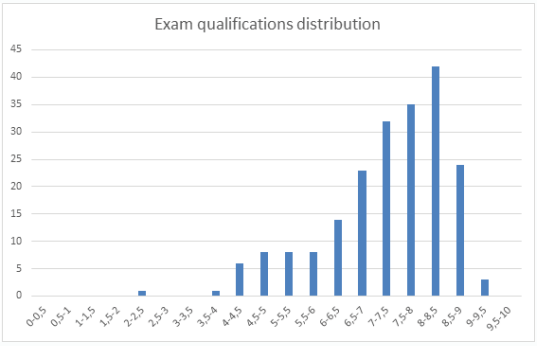

Figure 1: Results of the exam in a group (n=45) of "Introduction to Surgical Practice" the academic year 2019-2020. The 'x' axis shows the academic marks (1-4.9: failed; 5-6.9: approved; 7-9.9: remarkable; 9-9.9: excellent; 10: distinction/mention). The 'y' axis shows the number of students evaluated.

Results of the make-up evaluation

A second-chance exam for the course was organized for the students who failed the first examination during the pandemic's first wave. Every student who was identified with identical answering patterns in the MCQ test was invited to this evaluation. This second exam resulted in a similar distribution of marks.

After the final marks were made public, students confessed in an interview that most of them had made an exam group using WhatsApp. The only students penalised in the first exam were the ones who had an exact coincidence in their responses, but it seems the answer-sharing was generalised.

Discussion

The present study aimed to draft a consensus document between medical undergraduates and clinical lecturers about an alternative online evaluation during the first wave of COVID-19 in Spain due to mobility restrictions in exceptional circumstances. An online evaluation through a common agreement was undertaken under an evaluation contract from both sides which included the consensus reached during the debate. Exam results were analysed and compared, and some breaches in academic integrity were detected at the end of the course.

The first findings extracted during the debate referred to the academic community's psychological burden, which is in line with Spanish data published on 20 May 2020 in an epidemiological study led by the Carlos III Health Institute and the National Epidemiological Watch Network (RENAVE [10]). It showed that since the beginning of the COVID-19 health alert, 40,961 cases of infection among healthcare providers were reported in Spain, representing 24% of all COVID-19 cases in the country [9], including part of the educators

Table 3: Completely identical answering patterns identified.

Patterns	Test results
A	
Student 1	2 2 2 2 2 2 2 2 0 2 2 2 2 2 2 2 2 0 0 0 2 2 2 0 2 0 0 0 2 2 0 2 2 2 2 2 2 2 2 2
Student 2	2 2 2 2 2 2 2 2 0 2 2 2 2 2 2 2 2 0 0 0 2 2 2 0 2 0 0 0 2 2 0 2 2 2 2 2 2 2 2 2
B	
Student 3	2 2 2 2 2 2 0 2 0 2 2 2 2 2 2 0 2 0 2 0 2 2 0 2 2 2 2 2 2 2 0 2 2 2 2 2 2 2 2 2
Student 4	2 2 2 2 2 2 0 2 0 2 2 2 2 2 2 0 2 0 2 0 2 2 0 2 2 2 2 2 2 2 0 2 2 2 2 2 2 2 2 2
C	
Student 5	2 2 0 2 0 2 0 2 0 2 2 2 2 2 2 0 2 2 2 2 2 2 0 0 0 2 2 2 2 0 2 2 2 2 2 2 2 2 2 2
Student 6	2 2 0 2 0 2 0 2 0 2 2 2 2 2 2 0 2 2 2 2 2 2 0 0 0 2 2 2 2 0 2 2 2 2 2 2 2 2 2 2
Student 7	2 2 0 2 0 2 0 2 0 2 2 2 2 2 2 0 2 2 2 2 2 2 0 0 0 2 2 2 2 0 2 2 2 2 2 2 2 2 2 2

The numbers show the sequence of the 40 test questions of the exam. 0=wrong answer 1=not answered; 2=correct answer.

in medical faculties and medical students are doing practices in teaching hospitals. The digitalization and protection of face-to-face education were not enough developed in medical education within South-European hospitals. Furthermore, lecturers in medicine, who usually have clinical responsibilities in hospitals, faced increased work demands and abnormally high psychological pressure during the pandemic. In this context, the psychological distress students were facing, especially an increase of depression, followed by burnout and anxiety related to the evaluations [10-12]. Similarly, Nakhonstin-Ansari and colleagues [13] found that 38% of medical students had mild to severe anxiety during the pandemic, and 28% had depression, with higher levels in females and those who had experienced COVID-19 symptoms. Thus, the psychological burden during the beginning of the pandemic times seemed to be a global trend [14].

Minimising the impact on student's academic record has also been a concern and an objective for other medical faculties world-wide. To look for efficiency in evaluating and providing customized student feedback through electronic marking systems was also tested differently. For instance, the electronic Objective Structured Clinical Examination (eOSCE), which tests students' clinical skills and competences by tracing an itinerary covering different simulated clinical scenarios, was applied by Alshammari [15], an exception in pandemic times. Regarding the evaluation during these times in Spain, Gonzalez and colleagues [16] found contradictory results at the UAM compared with the present study. They showed a positive effect of the confinement on students' performance through a reward instead of an 'evaluation contract'. In our case, even if the results of the first exam were positive, we cannot know to what extent these results are falsified apart from the three patterns detected and the confession about cheating through WhatsApp.

Other international experiences show different scenarios that made rapid adaptations in higher education processes during the first wave. Students from the Aalborg University (Denmark [17]) and Harvard University (USA) continued to undergo practical training during the wave, using it as a chance to strengthen their mastery of such skills as mechanical ventilation and the use of Personal Protective Equipment (PPE). However, this approach would have been impossible in Spain because of the strict confinement and many infected people, as the disease hit first very dramatically to South Europe (especially Italy, Spain, and Portugal). Spanish medical education was not enough prepared to use online alternative educational methods. Their agents did not experience innovating so rapidly, even using advanced technologies in clinical practices (e.g., robotic colorectal surgery). In the UK, for instance, they adapted clinical placements to an online model designed to help students acquire clinical knowledge using Peyton's four-step approach: demonstration, deconstruction, comprehension, and execution/performance [18]. A model we know now can be adapted to our context, but after our debate remains uncertain, the translation of medical practical training only through online methods and current technologies in surgery practices. The need for a continuous evaluation strategy including a diverse set of options, online and face-to-face, together with the possibility to undergo practical training in hospitals using PPE outside the usual teaching periods, was one of the main conclusions.

On the other hand, based on the student survey results, approximately half of our students attended the webinar, which was voluntary, and almost 40% freely participated in the survey. The other half seem insufficiently engaged with the webinar or had misgivings regarding expressing their views of assessment issues to their lecturers, even when this was done anonymously and managed by researchers that were not their lecturers. In similar studies, the variability of responses and non-answered questions regarding the assessments or evaluations is common, especially in pandemic times [15]. It seems students usually expect fair assessment tools and welcome methods that provide meaningful feedback [19,20]. However, the pandemic produced an increase in workload for students and lecturers (e.g., UK [18]). Students were worried about the prospect of disappointing results appearing on their academic record, and there was no precedent event about how to tackle evaluations in this global situation [16,17].

During the pandemic, most Universities decided to apply online evaluation through video surveillance, for instance, with the Respondus monitor, an application that allows students to take exams at home while being recorded with a webcam. This software has shown less than 1% difference between the virtual and non-virtual testing groups preventing digital cheating [21]. In our case, only a third of students confirmed this was a fair option, as concerns emerged regarding privacy issues during the webinar and in the chat. Interestingly, there is not much educational research on online systems to prevent digital cheating, which affects universities worldwide during COVID-19. Half of our cohort expressed the urgent need to develop continuous evaluation with special protection of face-to-face lectures, especially practices in hospitals; online options could be part of the overall evaluation strategy but accompanied by face-to-face assessed activities and exams (i.e., well-designed evaluations can have positive effects on learning in medical students [20]).

Both students and lecturers expressed a need for agreement during the webinar, through an evaluation contract of mutual trust d to be a valid and effective tool for evaluating the course. However, findings show this agreement did not provide an adequate online evaluation framework using the traditional exam model in an online format. Seven exam submissions (3.4%) had identical submissions. Despite the promise to complete the evaluation individually at home without

digital cheating, most students could not avoid the temptation to cheat *via* WhatsApp® groups, as they privately confessed to the course's main lecturers once all the grades were published. The time limit to complete the evaluation provided enough time to share the answers with other students, despite having signed an evaluation contract. On the other hand, education needed to advance as medical students and young doctors were needed on the pandemic's front line, and the system needed to keep going. However, the experience has shed light on how to do it better next time.

Some challenges, opportunities, and changes have emerged in the findings achieved, which relate to the notion of liquid education [5] and liaise with specific recommendations:

- Evaluations were held in a critical environment with a high rate of depression and stress. We must evaluate students in a fair rather than a frustrating environment.
- We should seize this opportunity to improve by instituting new evaluation alternatives, such as open-book exams, peer evaluation. For instance, assessment activities must be multiple with a short answer time and at a final stage with less weight through an exam at the University.
- An agreement of trust between lecturers and students can be a solution if accompanied by other methods and non-intrusive technologies. The continuous evaluation strategy and future technologies can help with this challenge (e.g., virtual reality headset).
- Teachers and students must be able to provide feedback on evaluation activities and exams. Both agents are susceptible to be assessed and share co-responsibility to promote social liquid education to start changing educational processes (e.g., educational social networking sites).
- Lower grades are not an option to protect the quality, but evaluation can be conducted differently. Other options exist to protect students' grades from unexpected situations, which probably pass for providing more educational, diverse opportunities.
- The lecturers must establish clear rules before the evaluation with a level of flexibility in modelling the evaluation process during the semester, i.e., to start giving the responsibility of acquiring the students' learning objectives. However, the system needs to protect honest students and the academic integrity of the University.

Thus, the reflections regarding continuous liquid evaluation are the following, which are currently applied through 'Operemos' in the UAM teaching hospitals during this pandemic:

- The semesters can be strategically planned with a set of online and face-to-face assessment activities (ideally more than needed) susceptible to be formatively assessed to guarantee the measurement of knowledge and skills before the final exam, reducing the weight of the exam.
- The diversity of activities should promote different pedagogical opportunities, for instance, individual activities (by the student, e.g., a test, an OR observation, a clinical case) or in group activities (by peers or with the lecturer, e.g., workshops, debates, presentations).
- It can be assessed students' teaching capacity (i.e., if a student can effectively teach other students), and the assessment exercise can be bidirectional (i.e., between peers, students-lecturers) or multidirectional (i.e., student-peers-lecturers).
- The lecturer can propose to the student to achieve a set of learning objectives through an itinerary created by him/her, i.e., to personalize learning. The student chooses the multiple learning opportunities, which count for the evaluation (e.g., as an evaluation a la carte).
- Another more dynamic and open to change pedagogical approach to teach and assess flexibly is necessary. Using differently current technologies (Moodle) or creating new tools such as formal *vs.* informal platforms (e.g., Operemos *vs*. Medicalum).
- Universities are starting to change, but the human component still needs time to comprehend and behave differently to adapt us to the new ways of educating.

This study, however, has also its limitations. The sample size and sampling strategy limit generalization, but it was not the main methodological objective. This paper is an action-research in a narration style to inform about an experience in other Universities worldwide and have solutions as the proposed ones. Furthermore, the methods used are based on self-reported measures from natural narratives from students' and lecturers' voices and students' texts messages from the chat, a student' survey, and exam patterns of responses. However, it is an original applied mixed-methods research study that is unique and contributes to science on showing preliminary evidence of a fact happening in all Universities (i.e., digital cheating even using strategies such as randomizing MCQs responses in Moodle or applying proctoring methods such as Respondus) and sharing reflections on how difficulties remain and can be overcome if key factors are detected on time, methods are flexibly redesigned, educational agents trained and tools are differently used, or new platforms or other devices applied.

The present paper chronologically narrates how an environmental challenge such as the pandemic situation was an opportunity to join the academic community on reflecting on liquid education and continuous liquid evaluation. However, it failed in a proportion of students who digitally cheat. Two aspects have probably produced the outcome: (i) although there was a common intention to start creating a new evaluation format through a liquid education during the first wave, the current use of present technologies and the capacity of adaptation of both agents were not enough prepared to rapidly adapt to a new situation; (ii) the online exam had the same structure as the traditional one, with the common adjustments to make it fair through Moodle programming activities using all available options. This experience enhances the need for teaching in values, i.e., to connect academic and professional values when designing new liquid evaluation systems to make future medical professionals as good at behaving as they were during their years of training in University. This pandemic can be longer than expected, or other future external and environmental factors can produce similar situations (e.g., climate change), and credibility of training in pandemic times depends to a large extent on the recognition of the validity of the assessment made that involves students, lecturers, and professionals. Thus, we have addressed the need to protect the human factor.

Conclusion

Hence, we concluded at present, any online evaluation strategy based only on personal responsibility and the expectation that students will never cheat is bound to fail, at least for a small proportion of students. However, we can start applying Bauman's framework [4,5] to the current training and evaluation situation, considering traditional evaluation methods to be solid and the new evaluation model to be 'liquid'. This argument favouring 'liquid' evaluation should create a model that encourages continuous training, assessments, and final exams during unprecedented times, as part of the constant training process to ensure the system move on independently of any environmental factor. Liquid evaluation should foster student responsibility (e.g., training itinerary), creating an agreement of trust between educator and student, plus other aids to ensure honest practices are protected, while promoting a connection between academic and professional values (e.g., if students cheat on online exams, are they going to also cheat in the medical (digital) practice?). Another inference extracted from this experience is we need new educational tools and methods ready to flip at any time between environments with all agents ready to switch with safety and security measures to guarantee the system is fair for all sides and protecting face-to-face education and evaluation. Finally, as for the ethical aspects of liquid evaluation, the system should create a trustworthy and safe environment during the teaching and continuous evaluation, establishing a system in which lecturers can rapidly fix dishonest practices. These reflections should lay the groundwork to transform the classical evaluation model to the liquid continuous evaluation model and adjust University to the 21st century.

Disclosures

Compliance with ethical standards

All ethical standards were complied with whilst conducting this research.

Funding

The Quironsalud Award 2021, category 'Quality of Care' for the project 'Informed Consent 2.0.: study of the impact of shared decision-making mediated by an app between the cancer patient and the surgeon' developed by the Héctor Guadalajara Labajo, Olatz Lopez-Fernandez and Damián García-Olmo.

Conflict of interest

The authors declare that they have no conflicts of interest.

Ethical approval

Ethical approval was not required for this study.

Informed consent

Informed consent was obtained from all individual participants included in the study.

Acknowledgments

Professor Juan BotellaAusina, Faculty of Psychology, UAM, for the statistical analysis. AuxiliadoraGraciani Pérez-Regadera, student Vice-Dean of the UAM, for her participation in the student interviews. Professor Carlos Félix Sánchez Ferrer, Dean of the UAM. Dr Oliver Shaw for his English style correction. Jorge Llau García, Alfonso Cuenca Navarro, Guillermo Laguna Murcia, María Sasía García, Javier Jiménez López, Beatriz Atienza Carbonell, José Miguel García Abajo, and María Castro Vázquez, AutonomousUniversityof Madrid students, participated as panel lists in the webinar. Arroyo Sebastián, Antonio; Cañada Lahoz, Emma; Cascales, Pedro; De La Portilla, Fernando; Fernández Bermejo, Eva María; García Carmona, Antonio; Jarabo Tévar, Beatriz; Jiménez, Montiel; Lozano, Pilar; Ramírez, Jose Manuel; Peña Hidalgo, Adrián, participated in the online debate.

References

1. Murillo Pacheco J. Principios y propiedades de la evaluación. Bolivia: Sembrando conocimiento. 2017.
2. Struyven K, Dochy F, Janssens S. Students' perceptions about evaluation and assessment in higher education: A review. Assess Eval High Educ. 2005;30(4):325-41.
3. Urdan TC, Paris SG. Teachers' perceptions of standardized achievement tests. Educ Policy.1994;8(2):137-56.
4. Bauman Z. Liquid Modernity. Cambridge, UK: Polity Press. 2000.
5. Bauman Z. Education in liquid modernity. Review of education, pedagogy, and cultural studies. 2005;27(4):303-17.
6. Sahu P. Closure of universities due to Coronavirus Disease 2019 (COVID-19): impact on education and mental health of students and academic staff. Cureus. 2020;12(4):e7541.
7. Han H, Resch D, Kovach R. Educational Technology in Medical Education. Teach Learn Med. 2013;25 Suppl 1:S39-43.
8. Royal Decree 463/2020, of March 14, which declares the state of alarm for the management of the health crisis situation caused by COVID-19.
9. Barnett-Page E, Thomas J. Methods for the synthesis of qualitative research: a critical review. BMC Med Res Methodo. 2009;9(1):1-11.
10. Analysis of COVID-19 cases in health personnel notified to RENAVE until May 10 in Spain. Epidemiological Surveillance Network (RENAVE) 2020.
11. Enns MW, Cox BJ, Sareen J, Freeman P. Adaptive and maladaptive perfectionism in medical students: a longitudinal investigation. Med Educ. 2001;35(11):1034-42.
12. Sani I, Hamza Y, Chedid Y, Amalendran J, Hamza N. Understanding the consequence of COVID-19 on undergraduate medical education: Medical students' perspective. Ann Med Surg (Lond). 2020;58:117-9.
13. Nakhostin-Ansari A, Sherafati A, Aghajani F, Khonji MS, Aghajani R, Shahmansouri N. Depression and anxiety among Iranian Medical Students during COVID-19 pandemic. Iran J psychiatry. 2020;15(3):228-35.
14. Rotenstein LS, Ramos MA, Torre M, Segal JB, Peluso MJ, Guille C, et al. Prevalence of Depression, Depressive Symptoms, and Suicidal Ideation Among Medical Students: A Systematic Review and Meta-Analysis. JAMA. 2016;316(21):2214-36.
15. Alshammari E. Implementing eOSCE during COVID-19 lockdown. J Adv Pharmacy Edu Res. 2020;10(1):175.
16. Gonzalez T, de la Rubia MA, Hincz KP, Comas-Lopez M, Subirats L, Fort S, et al. Influence of COVID-19 confinement on students' performance in higher education. PLoS One. 2020;15(10):e0239490.
17. Rasmussen S, Sperling P, Poulsen MS, Emmersen J, Andersen S. Medical students for health-care staff shortages during the COVID-19 pandemic. Lancet. 2020;395(10234):e79-e80.
18. Khan H. An adaptation of Peyton's 4-stage approach to deliver clinical skills teaching remotely. MedEdPublish. 2020.
19. Duffield KE, Spencer JA. A survey of medical students' views about the purposes and fairness of assessment. Med Educ. 2002;36(9):879-86.
20. Velan GM, Jones P, McNeil HP, Kumar RK. Integrated online formative assessments in the biomedical sciences for medical students: benefits for learning. BMC Med Educ. 2008;8(1):52.
21. Gilbert L, Nunez-Argote L, Hunter K, Smith CT. Prevention of Digital Cheating with Respondus Monitor. In Proceedings of the Teaching with Technology Symposium; 2016 University of Arkansas for Medical Sciences, US. 2016.