

DID COVID-19 RESTRICTIONS IN HIGHER EDUCATION AFFECT STUDENTS' PERFORMANCES? EVIDENCE FROM AN ITALIAN UNIVERSITY

Maria Michela Dickson, Giuseppe Espa, Diego Giuliani

Department of Economics and Management, University of Trento, Trento, Italy

Rocco Micciolo

Department of Psychology and Cognitive Science, University of Trento, Trento, Italy

Abstract. *The 2020 year has been characterized by an unprecedented epidemic that has involved all countries in the world. The Covid-19 emergency is having a strong impact on all aspects of everyday life, including services to education. The shift to emergency remote teaching in the spring of 2020 has had a strong impact on training procedures, with the consequent readjustment both in methods of teaching and in learning approaches. A natural research question concerns the impact on students' performance, as measured by exam scores, of the face-to-face teaching suspension. In order to understand the effects of the sudden transition to emergency remote teaching, in the present work a thorough comparison has been carried out between the outcomes of the exams taken by all students of a medium-sized University in Italy before and after the occurrence of Covid-19 pandemic. Results may be helpful in understanding advantages and limitations of distance learning, which seems to have become a new standard in the present and near future.*

Keywords: *Covid-19; higher education; distance learning; students' performance.*

1. INTRODUCTION

The COVID-19 pandemic is changing several aspects of life, not least everything related to education. The shutdown of social activities all over the world has implied the fast transition to online learning at all educational level, unearthing disparities among countries in terms, first of all, of internet access and technological tools availability (OECD, 2020). Moreover, another issue to consider concerns the effectiveness of online learning compared to traditional classrooms (Robinson and Hullinger, 2008; Means et al., 2014), which may vary amongst age groups (Simonds and Brock, 2014) and educational level.

Distance learning has a long history that spans two centuries (Spector et al., 2008). Clearly, a difference exists among online learning and emergency remote teaching: the former includes a variety of well-planned tools, such as, among others, distributed learning, blended learning and mobile learning, while the latter consists of online courses organized in response to a sudden and unpredictable

¹ Corresponding author: Maria Michela Dickson, email: mariamichela.dickson@unitn.it

crisis or disaster. Online education can lead to very efficient results if it is the product of a careful instructional design that considers pros and cons of distance learning. This planning is absent during emergency migration to online, in which face-to-face teaching is suspended until the emergency is ended and the main aim is to provide a necessary temporary access to education. In such situations, feasible solutions consist of blended and hybrid learning, as experienced in many crisis and war situations (Davies and Bentroto, 2011).

The emergency situation in which many European countries have found themselves since the first months of 2020 does not differ from those aforementioned. The first Western area suffering from the Coronavirus outbreak has been Italy, where in one weekend all aspects of life, included education, significantly changed. As an attempt to limit the diffusion of the virus, the Italian government issued on 4 March 2020 a Ministerial decree (DPCM decree) which, amongst other measures, ordered the closure of all schools and universities nationwide starting from 5 March. As a consequence, in order to ensure the continuity of didactics and education, both schools and universities had to suddenly adjust to emergency remote teaching. This shift was not free from problems and consequences, but also from opportunities, especially in higher education. In particular, as it is also pointed out by Gonzalez et al. (2020), it is of particular interest to evaluate the impact that emergency remote teaching may have on students' performance at university exams, which represent the most immediate, despite being imperfect, acknowledgment of the outcome of teaching activities. Indeed, while the literature has that students' performances tend to be similar when the same professor is teaching face-to-face and online classes (Stack, 2015) under normal conditions, it is noteworthy to understand if this remains valid also in emergency situations.

The first wave of the Covid-19 outbreak in Italy and the consequent suspension of face-to-face classes coincided with the second semester of the academic year 2019/2020. Therefore, since the beginning of March 2020 both teaching activities and the associated exams have been conducted according to emergency remote procedures, which have deeply changed the usual approaches and procedures followed by lecturers before the outbreak. In particular, we consider the students at the University of Trento, Italy, between 2019 and 2020, and we conduct a comparison between pre- and post-Covid-19 exam sessions. This athenaeum has a medium size, with 16,868 students across 90 undergraduate, graduate, postgraduate and PhD courses. As in many other universities (see, e.g., George, 2020), at the University of Trento many exams that were initially planned to be written have been held as online oral exams or

in a mixed oral-written online form, while other exams have been held as online written exams under some form of remote video supervision.

The aim of the present work is threefold. First, it investigates if the sudden shift to emergency remote teaching caused a reduction in students' attendance, which should lead to the observation that a non-negligible share of students, who could not adapt well to the change, decided to postpone some exams to subsequent exam sessions, slowing down their studies. Second, it assesses if potential changes in exam participation rates exist due to the remote form of examination, which may be positive (students perceive online exams as easier) or negative (students find more difficult to take online exams because of the complexity of the procedures of remote video supervision). Third, it finds evidence of potential changes in average students' grades due to both students' study behavior and lecturers' choices about modifications of the evaluation approach and of the level of difficulty.

In order to place the analysis that will follow, and the evidence that will emerge, into proper context, in Section 2 we provide details about the Italian university system and some of its peculiarities. In Section 3 an in-depth comparison in terms of performances of students among the second semester of 2019 and of 2020 is conducted, while in Section 4 the attention is focused on the existing variability across study programmes. The empirical results are discussed in Section 5 and Section 6 concludes.

2. THE ITALIAN UNIVERSITY SYSTEM

In accordance with dispositions contained in the Bologna Declaration (2009), the Italian university system is structured mainly in three cycles. The first cycle is represented by the *Laurea Triennale*, corresponding to *Bachelor's degree* in the European university system, which is a three-year programme of study that is awarded to students who gain at least 180 ECTS (European Credit Transfer System). The *Laurea Triennale* gives access to the second cycle study programmes. The prominent second cycle study programme is the *Laurea Magistrale*, corresponding to *Master's degree* in the European university system, which is a two-year programme of study that is awarded to students who gain at least 120 ECTS. Some programmes (such as those in construction engineering and law) that are defined *Corsi a Ciclo unico*, merge the first and second cycles into a single cycle degree programme. They last five years and award a *Laurea Magistrale Quinquennale* to students who gain at least 300 ECTS. The *Laurea Magistrale* and the *Laurea Magistrale Quinquennale* give access to the third cycle study, which mainly consist on PhD programmes. Our analysis focuses on

all the first, second and single cycles programmes offered by the University of Trento.

At Italian universities, the academic year is organized into two semesters: typically, the first one runs from September (or October) to January (or February) and the second one spans basically between February (or March) and June (or July). The actual starting and concluding dates of semesters vary across universities, but commonly each semester has a duration of about 20 weeks, which comprises a first teaching period of about 14 weeks and a subsequent exam session of about 6 weeks. Hereinafter, the regular exam sessions that follow the end of the teaching sessions will be called “winter exam session” and “summer exam session”. Normally, students can take exams in both winter and summer sessions once the semester lectures included in their study plan have concluded. It means that students who do not take (or fail at) exams of the first semester’s courses in the subsequent winter exam session have the opportunity to retake them during the next summer exam session. At the same time, students who do not take (or fail at) exams of the second semester’s courses in the subsequent summer exam session have the opportunity to retake them during the next winter exam session of the following year. In addition, students can benefit of a 3- or 4-weeks recovery exam session, which is usually held in September.

Depending on lecturers’ choices and the orientation of the different study programmes, exams can be oral, written, or both. According to a national norm, exam grades must follow a 30-point scale. An exam is considered passed if the grade is at least 18 out of 30.

3. AN EMPIRICAL COMPARISON BETWEEN EXAM SESSIONS BEFORE AND DURING COVID-19 PANDEMIC

The dataset used to conduct the analyses presented in this paper concerns the students, and their performances, at the University of Trento. The Institute offers 23 first cycle study programmes and 33 second cycle study programmes, plus two single cycle study programmes, postgraduate courses, PhD programmes and other programmes joined with other athenaeums. In Appendix A is reported the list of study programmes to which data used in the following belong.

The first empirical analysis comprises three parts concerning different perspectives of the phenomena of interest. Firstly, we make a general comparison between the overall exam result trend of the summer exam session of 2019 (before the pandemic) and that of the summer exam session of 2020 (during the pandemic). Secondly, by focusing only on the subset of exams for which the lecturer is the same across the second semesters of 2019 and of 2020, we perform a matched-pairs analysis based on the comparison within the same *lecturer/exam*

pairs across the two academic years. This approach allows us to control for confounding factors that could affect the results, such as a change of lecturer or a change of syllabus between 2019 and 2020, and isolate the effect of the shift to distance learning. Thirdly, we perform the same analysis of the matched *lecturer/exam* pairs while focusing on the comparison between the exams of the courses that were held during the first semesters, instead of the second, of the two academic years. This last part of the analysis lets us isolate the “pure” effect of the exam since it focuses on exams related to teaching activities that were held before the Covid-19 emergency in the traditional face-to-face forms.

3.1 AN OVERALL COMPARISON BETWEEN THE SUMMER EXAM SESSIONS OF 2019 AND OF 2020

The first step of the analysis focuses on the 6,084 exams taken by the students of the first, second and single cycle programmes of the University of Trento. Data are related to the summer exam sessions of 2019 (3,002 exams) and of 2020 (3,082 exams).

The forms of examination have changed substantially between the two academic years; indeed, the number of oral exams has increased by 28% (from 1,670 to 2,132), while the number of written exams has decreased by the same proportion (from 1,182 to 846), and the exams that comprise both a written and an oral test decreased by 31% (from 150 to 105).

The total number of students who took exams in the summer exam sessions of 2019 and of 2020 are 39,663 and 39,333, respectively. The proportion of exams with a positive outcome is equal to 0.913 (2019) and to 0.921 (2020), with a relative risk (Porta, 2008) equal to 1.008. Looking at the average grade of the exams, they are equal to 26.31 in 2019 and to 26.57 in 2020. It is clearly showed that there are no relevant differences between the two exam sessions. First of all, the number of students participating in the exam session decreased by only 330 units (which correspond to less than one percent). Secondly, the proportion of students who passed the exams is essentially the same. Moreover, the increase in terms of average grade (namely 0.26) is negligible considering the 30-point grading scale.

These results provide strong evidence that the impact of the switch to emergency remote teaching and examination on students' performances has been substantially negligible. However, in order to verify that the impact is not masked by confounding factors, the analysis has been repeated stratifying the exams according to some relevant dimensions. Table 1 reports the results stratified by, respectively, cycle programme, form of examination and corresponding teaching semester, one at a time. We do not detect relevant differences between 2019 and 2020 across any of the stratification variables – if we exclude the one concerning

Tab. 1: A stratified comparison between the summer exam sessions of 2019 and of 2020 of the University of Trento by cycle programme, form of examination and corresponding teaching semester

Stratification by	Number of students taking the exams			Proportion of passed exams			Average grade		
	2019	2020	Difference	2019	2020	Relative Risk	2019	2020	Difference
Cycle programme									
First	23,920	23,611	-309	0.889	0.900	1.012	25.77	26.04	-0.27
Second	8,351	8,625	+274	0.966	0.968	1.002	27.51	27.79	-0.27
Single	7,392	7,097	-295	0.933	0.933	1.000	26.61	26.77	-0.15
Examination									
Written	16,471	21,807	+5,336	0.967	0.970	1.003	27.10	27.15	-0.05
Oral	21,465	16,606	-4,859	0.872	0.858	0.984	25.68	25.70	-0.02
Mixed	1,727	920	-807	0.921	0.900	0.977	26.15	27.05	-0.89
Teaching semester									
First	3,574	3,422	-152	0.924	0.940	1.018	26.82	26.69	-0.13
Second	7,324	7,355	+31	0.845	0.856	1.013	25.39	25.74	-0.34
Whole year	28,765	28,556	-209	0.930	0.935	1.006	26.48	26.76	-0.28

the number of written and oral exams, which is due to the fact that lecturers have been mostly invited to shift to oral exams.

Regarding students' performance – evaluated both in terms of relative risk and difference in the average grade – the two considered summer exam sessions are essentially identical. This is evident, in particular, with respect to the teaching semester. Indeed, the courses held in the second semester are those that were expected to be mostly affected by the emergency health situation since their lectures were suddenly and necessarily held online, along with the associated problems in reprogramming entirely the courses and the difficulties encountered in using on-line platforms. Quite surprisingly, however, according to the results of Table 1, this does not seem to be the case: the differences between 2019 and 2020 are indeed irrelevant despite the semester in which the courses were held. This result was highly unexpected and suggest that: (i) lecturers did their best to rapidly adapt teaching forms and materials to the emergency situation and (ii) this generation of students, which possess a high computer literacy, positively responded to the use of informatic tools in the learning process.

Not surprisingly, there are instead relevant differences across the two academic years within the strata. In particular, the proportion of passed exams, as well as the average grade, are higher for second cycle programmes than first cycle programmes; they are higher for oral examinations than written examinations; and they are higher for exams that refer to courses held in the second semester than in the first semester. We could argue that there was a double side effect that contributed to this result: on the one hand, lecturers may have been more sympathetic towards students at oral examination and, the other hand, second cycle programmes' students may have been more reactive to on-line examination thanks to their larger experience.

3.2 MATCHED-PAIRS ANALYSIS FOR THE SECOND SEMESTER

The analysis that follows further reinforces the evidence of a strong stability of students' performances across 2019 and 2020. It focuses on a selected subset of exams of the courses held during the second semester of both 2019 and 2020. The selection is based on forming "pairs" of teaching activity and corresponding lecturer and then matching them across the two academic years. In other words, we built a treatment-control scenario where the occurrence of Covid-19, and the consequent shift to distance learning, represents the treatment, pre-Covid-19 exams form the control group and post-Covid-19 exams identify the treatment group. The control and treatment groups are made comparable with respect to confounding factors by individually pairing the exams of the two groups with respect to course, lecturer and semester. Therefore, each exam in the treatment group has a "twin" in the control group that is characterized by the same course, lecturer and semester. We have thus performed a matched-pairs analysis (Agresti, 2018, chapter 9), in which the differences between pre- and post-Covid-19 exams that depend on teaching semester and lecturer are filtered out, that is they are controlled for by the matching procedure.

The number of all distinct lecturer/exam pairs that occur in the two considered summer exam sessions is 1,080 (812 in 2019 and 831 in 2020). The pairs that occur in both semesters are 563 (that is around 2/3 of pairs in 2019); the pairs that occur only in 2019 are 249; those that occur only in 2020 are 268.

The matched-pairs analysis is therefore focused on the 563 exams of the courses in which the lecturers have remained the same in 2019 and 2020. In this selected subset, it is reasonable to expect that the differences across the two academic years are more marked with respect to those obtained considering all the exams, especially in terms of average grade.

At whole-university level, the result obtained with the selected matched-pairs sample confirms the evidence obtained by analyzing the whole set of exams. In fact, the matched-pairs analysis highlights that the proportion of passed exams is equal to 0.933 in 2019 and to 0.936 in 2020, with a relative risk equal to 1.004. The average grades in the two considered exam sessions go from 26.50 in 2019 to 26.76 in 2020, with an increase of 0.26 on a 30-point grading scale. These results are almost equal to those reported in the previous section, where the analysis does not consider matched-pairs. The unexpected evidence that the performance at the exams of the students of the University of Trento is not changed across 2019 and 2020 is therefore confirmed.

To further verify the solidity of this evidence, in Table 2 we report the results about the matched-pairs disaggregated by cycle programme and types of examination.

Tab. 2: A comparison between the summer exam sessions of 2019 and of 2020 of the University of Trento by cycle programme and type of examination (based on the second semester's matched-pairs)

	Number of students		Proportion of passed exams			Average grade		
	2019	2020	2019	2020	Relative risk	2019	2020	Difference
Cycle programme								
First	13,442	13,125	0.910	0.917	1.008	25.94	26.32	+0.37
Second	4,508	4,506	0.972	0.967	0.995	27.63	27.79	+0.16
Single	3,943	3,796	0.965	0.968	1.003	27.06	27.09	+0.03
Examination								
Oral to Oral	8,535	8,817	0.975	0.980	1.005	27.25	27.36	+0.11
Written to Written	9,051	9,082	0.886	0.884	0.998	25.66	25.90	+0.24
Written to Oral	2,741	2,361	0.941	0.960	1.020	26.57	27.29	+0.72
Oral to Written	449	468	0.982	0.908	0.925	27.11	26.90	-0.21

The evidence remains the same even within cycle programmes: students' performance has not been noticeably affected by the sudden change of teaching or examination approaches.

Concerning the type of examination, if we exclude the exams that were held in a mixed oral-written form in 2019 or 2020 (a total of 29 exams), we observe 303 out of 534 (56.6%) exams that were oral in both sessions, 138 out of 534 (25.8%) that were written in both sessions, 80 out of 534 (15.0%) that moved from the written to the oral form, and finally 14 out of 534 (2.6%) that moved from the oral to the written form. From the conducted analysis, some useful evidences emerge:

- The number of students who took oral exams and the number of students who took written exams are very similar; however, the number of oral exams is greater than the number of written exams. This implies that, in general, the courses with a written exam tend to have a higher number of students.
- The number of students who took the exams that were written in 2019 and become oral in 2020 is similar across the two exam sessions (2,741 versus 2,361). Therefore, the change from the written to the oral form of examination does not seem to have affected, neither positively nor negatively, exam participation. The same can be said for the opposite change, from the written to the oral form of examination, but with quite lower numbers (449 versus 468).
- Both the proportions of passed exams and the average grades are higher for the oral form.
- The exams that did not change the form of examination show approximately identical results, both in terms of proportions of positive results and average grades, across the two academic years.

- The change from the written to the oral form of examination mainly pertain those exams which already had a relatively higher proportion of positive outcomes than the exams that remained written. Indeed, in 2019, the proportion of positive outcomes for the exams that remained written is 88.6%, while the same proportion is 94.1% for the exams that moved from the written to the oral form.
- For the small subset of exams that moved from the oral to the written form of examination, we observe a slight decrease of the proportion of passed exams across the two academic years (from 0.98 to 0.91), while the average grade remains essentially the same (27.1 versus 26.9).

In conclusion, even if we disaggregate the matched-pairs by form of examination, we do not observe any substantial difference in terms of performance between the two periods – while there are expected non-negligible differences within periods and across the forms of examination. The only variation that has some relevance regards the change from the oral to the written form, but involves only a small number of students.

3.3 MATCHED-PAIRS ANALYSIS FOR THE FIRST SEMESTER

In order to better assess the substantial meaning of the obtained results, the matched-pairs analysis is here repeated considering a different subset of exams. In particular, in this case, we focus on the exams taken in the summer exam session of both 2019 and 2020 but that refer to courses held in the corresponding first semesters. Even in this analysis, the units of observation are represented by matched-pairs of teaching activity and corresponding lecturer in which the lecturer remained the same across the two academic years. Unlike the previous analysis, in this subset the effect of the shift to emergency remote teaching is absent since only the examination moved to online procedures.

The number of all distinct lecturer/exam pairs that refer to teaching activities held in the first semesters is 830 (649 in the first semester of 2019 and 652 in the first semester of 2020). The pairs that occur in both academic years are 470, that is slightly less than 3/4 of pairs in 2019; the pairs that occur only in 2019 are 178; those that occur only in 2020 are 181. The analysis focuses on the 470 exams in which the lecturer is the same across the two academic years.

Table 3 summarizes the results of both this analysis and the previous two. In particular, the first row refers to the matched-pairs concerning courses held in the first semester, while the second row regards the matched-pairs concerning courses held in the second semester. The results can be undoubtedly considered as equivalent to those obtained considering all the exams. However, the first one – concerning first semester's courses – should be affected by the shift to remote

examination only, while the second one – concerning second semester’s courses – should be also affected by the shift to remote teaching.

Tab. 3: A comparison between the second summer exam sessions of 2019 and of 2020 of the University of Trento (based, respectively, on the first semester’s matched-pairs, on the second semester’s matched-pairs and on all exams)

Sample	Number of students		Proportion of passed exams			Average grade		
	2019	2020	2019	2020	Relative risk	2019	2020	Difference
1st Semester	6,015	5,805	0.841	0.849	1.010	25.28	25.60	+0.32
Matched-pairs								
2nd Semester	21,893	21,427	0.933	0.936	1.004	26.50	26.76	+0.26
Matched-pairs								

Table 4 shows the results disaggregated by cycle programme. If we compare them with the analogous results for the matched-pairs of the first semester’s courses we can see that the most striking (and obvious) difference between the two semesters concerns the distinctly smaller number of students taking the exams.

Tab. 4: A comparison between the summer exam sessions of 2019 and of 2020 of the University of Trento by cycle programme (based, respectively, on the first semester’s matched-pairs, on the second semester’s matched-pairs and on all exams)

Cycle programme	Number of students		Proportion of passed exams			Average grade		
	2019	2020	2019	2020	Relative risk	2019	2020	Difference
First	3471	3229	0.784	0.808	1.031	24.51	24.92	+0.42
Second	1434	1352	0.956	0.964	1.008	26.77	27.05	+0.28
Single	1110	1224	0.873	0.831	0.952	25.37	25.47	+0.10

Even the proportion of passed exams, if we exclude the second cycle programmes, is non-surprisingly smaller. The two summaries are however very similar with respect to the comparison between 2019 and 2020, both in terms of proportion of passed exams and average grade.

4. COMPARING THE VARIABILITY ACROSS STUDY PROGRAMMES BETWEEN EXAM SESSIONS BEFORE AND DURING COVID-19 PANDEMIC

The sudden shift to emergency remote teaching and examination, even if it did not affect the average exam performance, may have had an impact on the variability of exam performance. For example, study programmes characterized by laboratory-based courses, which normally require the physical presence of students, may have been negatively affected by the forced change to online teaching; while study programmes with many crowded courses may have paradoxically benefited from the new situation. In other words, the average equivalence between the exam sessions of 2019 and of 2020 may be the result of the compensation between negative and positive effects. If that would be the case, we should observe an increased variability of exam performance across study programmes in 2020.

4.1 PERFORMED MODELS

In order to evaluate this possibility, we have fitted some mixed-effects models (Snijders and Bosker, 1999; Deleeuw et al., 2007) treating study programmes as random effects. In particular, for the summer exam session of 2019 and of 2020, we have fitted different models considering two distinct response variables (i.e. average grade and proportion of passed exams) evaluated separately for two different cycles (i.e. first cycle and second cycle). Each model was fitted on the exam results grouped by study programme including a group-level random intercept.

More formally, for the models treating the average grade as the dependent variable we used the following linear specification:

$$y_{ij} = \mu + \delta_j + \varepsilon_{ij} \quad (1)$$

where y_{ij} represents the average grade of the i – th exam belonging to the j – th study programme, μ denotes a fixed effect, and δ_j is a random effect referring to the j – th study programme. Both in 2019 and in 2020, the model for the first cycle is characterized by 23 study programmes, and hence includes 23 random effects; while the model for the second cycle is characterized by 33 study programmes, and hence includes 33 random effects (see Appendix A for the list of study programmes). This analysis has not been performed for single cycle study programmes due to the lower number of observations.

In this framework, δ_j can be interpreted as the random deviation of the j – th study programme' average grade from the overall average grade μ , thus

implying that the average grade of the $j - th$ study programme is $\mu + \delta_j$ (in this respect δ_j is often called a “random intercept”).

The variance of the random intercept δ_j is σ_δ^2 , while the variance of the residual ε_{ij} is σ^2 . Under the assumption that ε_{ij} and δ_j are independent, the variance of the responses y_{ij} is given by $\sigma_\delta^2 + \sigma^2$ (i.e. the sum of the between and within study programme variances). Therefore, the proportion of the total variance is given by

$$\frac{\sigma_\delta^2}{\sigma_\delta^2 + \sigma^2},$$

which is called the (unconditional) intraclass correlation coefficient (ICC).

Under the further assumption that the error ε_{ij} and the random effects δ_j are independent and identically distributed normal random variables, with zero mean and variances σ^2 and σ_δ^2 , respectively, the main interest is in estimating σ_δ^2 and in testing the null hypothesis that $\sigma_\delta^2 = 0$, i.e. no variability between study programmes (i.e. there is no random intercept in the model: $\delta_j = 0$).

The grouping structure of exams into study programmes may imply the violation of the independence assumption underlying the standard OLS estimator; indeed, observations within each study programme are more likely to be similar to each other than observations belonging to different study programmes. Mixed-effects models, on the other hand, allow to explicitly control for the dependence within groups via random effects parameters, exploiting the information about the grouping structure of data.

Moreover, since the proportion of passed exams is a quantity limited between 0 and 1, for the models treating this variable as the dependent variable we used the following logistic specification (Agresti, 2003; Rabe-Hesketh and Skrondal, 2012),

$$\text{logit}(p_{ij}) = \beta + \delta_j, \quad (2)$$

where p_{ij} represents the proportion of positive outcomes of the $i - th$ exam belonging to the $j - th$ study programme, β denotes a fixed effect representing the odds of passing an exam, and δ_j is a random effect referring to the $j - th$ study programme. As in the linear model (1), the variance of the random intercept δ_j is σ_δ^2 and has a similar interpretation.

For each fitted mixed effect model, the significance of the random effect at the study programme level ($H_0: \sigma_\delta^2 = 0$), was assessed employing the likelihood-ratio test as described in Rabe-Hesketh and Skrondal (2012).

4.2 RESULTS

Table 5 shows the estimates of parameters μ and σ_{δ}^2 , the 95% confidence intervals of σ_{δ}^2 , the corresponding significance test and the estimated ICC for the linear mixed effects models of the average grade. Even if the variability between study programmes accounts only for a small percentage of the total variability, for all models, the test for the random effect is always highly significant, thus providing strong evidence of heterogeneity in exam performance across study programmes.

Assuming normality, employing the estimates of the standard deviation of the intercept, it is possible to form intervals within which 95% of the study programmes random intercepts are expected (these intervals represent ranges within which 95% of the realizations of a random variable are expected to lie and should not be confused with confidence intervals). In particular, these intervals are derived from a Normal distribution with mean equal to the estimated value of μ and standard deviation equal to the estimated value of σ_{δ} and help in assessing the degree of heterogeneity across study programmes. For first cycle study programme these intervals are (24.4, 27.3) in 2019 and (24.8, 27.8) in 2020; for second cycle study programme these intervals are (25.3, 29.6) in 2019 and (26.4, 29.1) in 2020.

Tab. 5: Estimates of the mixed-effects linear regression models of the average grade

	μ	σ_{δ}^2	95% confidence interval of σ_{δ}^2	Significance test for σ_{δ}^2	ICC
2019					
First Cycle	25.9	0.550	(0.248; 1.220)	33.5 (p-value < 0.001)	0.098
Second Cycle	27.4	1.202	(0.651; 2.222)	81.1 (p-value < 0.001)	0.228
2020					
First Cycle	26.3	0.565	(0.264; 1.208)	42.8 (p-value < 0.001)	0.112
Second Cycle	27.8	0.475	(0.241; 0.933)	44.1 (p-value < 0.001)	0.124

The estimates of the mixed-effects logistic regression models of the proportion of passed exams provide similar evidence (Table 6). The variances of the random intercepts were very similar in the two cycles as well as in the exam sessions considered. The estimates of the fixed effect β shown in Table 6 refer to the estimated odds of passing an exam and can be converted in estimated probability $\hat{p}_{ij}$ according to model (2) employing the formula $\exp^{\beta}/(1 + \exp^{\beta})$. These estimates $\hat{p}_{ij}$ for the first cycle study programme are 0.930 both in 2019 and in 2020, while for the second cycle study programme are 0.979 and 0.985. Therefore, on average, the proportion of positive outcomes for the exams of the first cycle study programmes remains substantially the same between the pre-

Tab. 6: Estimates of the mixed-effects logistic regression models of the proportion of passed exams

	β	σ_{δ}^2	95% confidence interval of σ_{δ}^2	Significance test for σ_{δ}^2	ICC
2019					
First Cycle	2.588	0.929	(0.673; 1.282)	743.2 (p-value < 0.001)	0.208
Second Cycle	3.839	1.019	(0.715; 1.451)	108.8 (p-value < 0.001)	0.240
2020					
First Cycle	2.593	0.806	(0.586; 1.107)	633.0 (p-value < 0.001)	0.165
Second Cycle	4.181	1.407	(0.991; 1.997)	207.4 (p-value < 0.001)	0.376

and post-Covid-19 periods. On the other hand, among the second cycle study programmes, the average proportion of passed exams slightly increases between the two periods.

The estimated expected distributions of the study programme-level random effects are shown in Figure 1. For the first cycle programmes' models of the average grade the estimated standard deviations of random effects for 2019 and for 2020 were quite similar and not significantly different (p -value = 0.933) according to a randomization test (Edgington and Onghena, 2007; Good, 2013). On the other hand, if we consider the second cycle programmes' models, we observe a significant reduction of the variability between 2019 and 2020 (with a randomization test that yielded a p -value of 0.007).

For the models of the proportion of passed exams, the estimated standard deviations for 2019 and for 2020 were quite similar among first cycle programmes (and not significantly different according to a randomization test: p -value = 0.565). On the other hand, among second cycle programmes, we observe an increase in the variability between 2019 and 2020 though it is not significant according to a randomization test (p -value = 0.182).

We can therefore conclude that the change of the forms of teaching and examination due to the health emergency situation did not relevantly increase the heterogeneity of students' performance at exams across study programmes, both in terms of proportion of positive outcomes and average grade. Furthermore, in the case of first cycle programmes, the variability of the average grade, instead of increasing, has been significantly reduced.

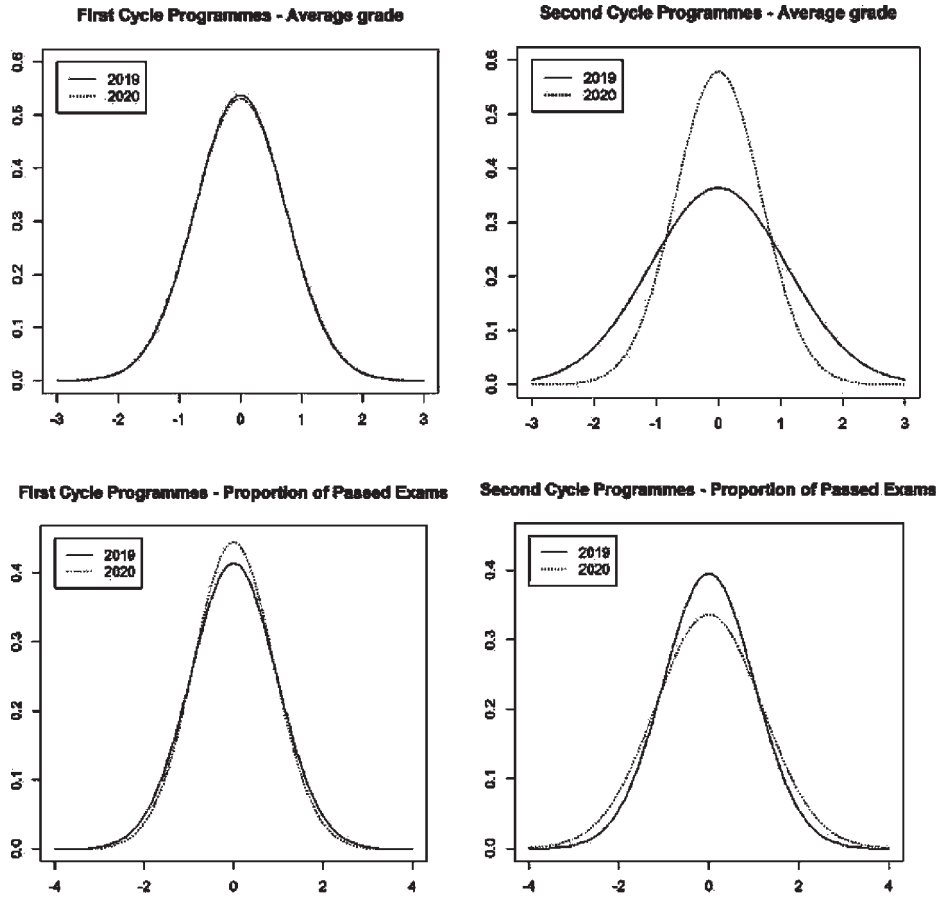


Fig. 1: Estimated expected distributions of the study programme-level random effects of the four mixed-effects regression models. x-axis: study programme-level deviations from the global intercept; y-axis: normal density

In order to better assess how the variability of students' performance has changed across the two periods, and identify which study programmes have been characterized by relevant changes, it is possible to predict the value of individual study programme-level random intercepts δ_j treating the parameter estimates μ , σ_δ^2 , σ^2 as true parameter values and employing *empirical Bayes prediction* (Rabe-Hesketh and Skrondal, 2012). The prior distribution is the normal distribution for the random intercept with zero mean and variance $\hat{\sigma}_\delta^2$, while the likelihood for the j -th study programme can be found employing maximum likelihood estimation.

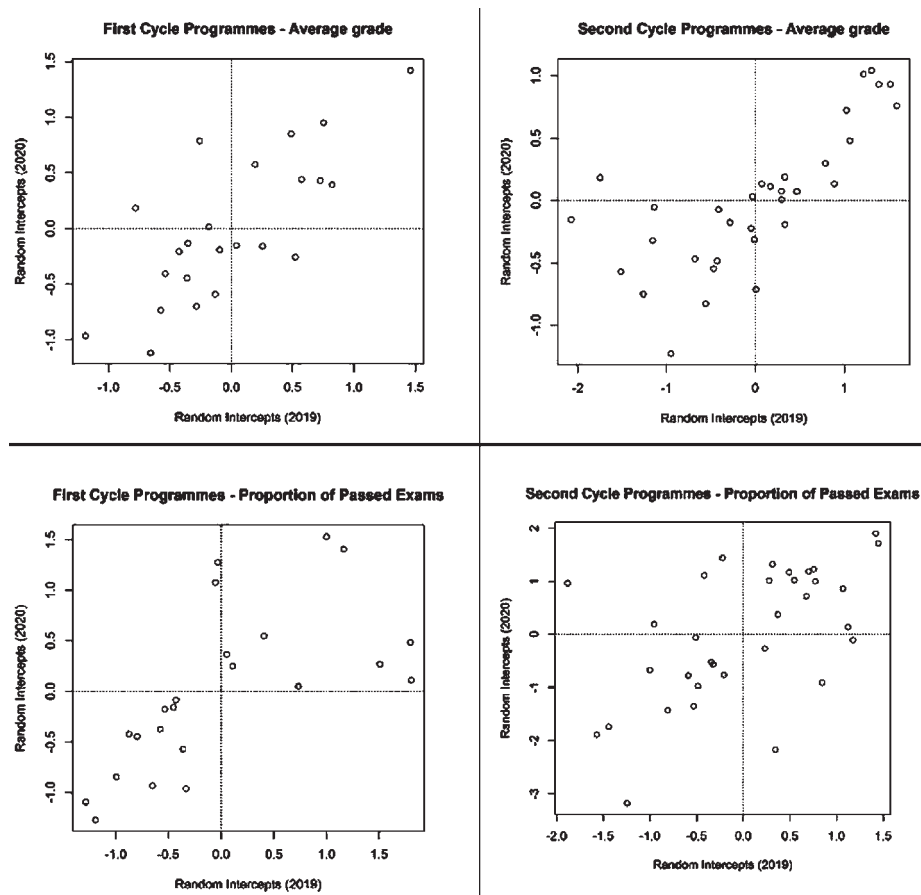


Fig. 2: Comparison between the predicted study programme-level intercepts in 2019 and in 2020 across the four estimated mixed-effects models

The estimated predicted values of the random intercepts for the different study programmes in 2019 and in 2020, separately for the two cycles and the two considered dependent variables, are shown in Figure 2. In general, there is high concordance between the performance of study programmes in 2019 and in 2020. With respect to the average score, about 74% of first cycle study programmes and 88% of second cycle study programmes maintained the same above (or below) the average performance. Similarly, with respect to the proportion of passed exams, these percentages are 91% and 76%, respectively. These estimates are naturally based on the assumption that, in each of the four models, the study programme-level intercepts follow a normal distribution. This assumption has been checked through the Q-Q plots for the corresponding empirical Bayes predictions, which did not provide evidence of violations of normality.

5. DISCUSSION AND CONCLUSIONS

The present work has been focused on the analysis of the impact of the shift to emergency remote teaching on student's grades. The data used refer to students of all the first, second and single cycle programmes offered by the University of Trento, Italy. In the first part of the paper, we performed a comparison in terms of number of taken exams, stratified by cycle programme, form of examination and corresponding teaching semester, among the second summer exam sessions of 2019 and of 2020. Then, a matched-pairs analysis has been performed, controlling for lecturer and semester, about the courses held in the first and second semester, with exams taken only in the summer exam session of 2019 and of 2020. Lastly, we compare the variability across study programmes among exam sessions before and during the health emergency. The obtained results reveal a somewhat surprising homogeneity in students' performances across the exam sessions before and during the Covid-19 pandemic. The worries that have motivated these analyses and, in particular, the concern that the emergency remote teaching could negatively affect the natural progression of students' careers should be scaled down. The same can be said about the opposite concern that online exams may promote opportunistic behaviors by students and, in turn, could cause an increase in the exams' outcomes.

This encouraging evidence, however, should be considered with some cautions. First of all, it may be possible that the observed results are only valid for the share of students who attended regularly the online courses. The results about the comparison between the matched-pairs of the first semester's courses should reassure in this respect. However, it is not clear yet which is the effect of a protract lack of a direct contact with the lecturers on the students that do not attend, or attend occasionally the courses. For this reason, it is important to maintain a constant monitoring of the situation.

Secondly, it should be considered that the emerged evidence could be the result of the compensation of different systematic effects. A hint in this direction is given by the high number of exams which changed the form of examination from written to oral. It is common knowledge that oral exams tend to lead to better outcomes with respect to written exams. Another confounding effect may arise if the negative outcomes of oral exams are not registered. Indeed, in such a case, the relative risk would decrease, thus compensating the effect of emergency remote teaching.

Further systematic compensation effects could be due to the fact that online exams are taken under less controlled conditions, to the fact that students are less stressed when take the exams at home, or because lecturers tend to be less strict in grading, in order to compensate for the difficulties that the students met when

attending the course in this new form. Some of these compensation effects could be explored with further surveys and data collections.

Despite these caveats, the results suggest that, in general, students and lecturers in Italy have reacted to the conditions imposed by the emergency health situation in a positive way and without causing relevant effects on the progression of students' careers and grades. Besides, the most important caution should be addressed to the comprehensive evaluation of the exam as an integral component of the educational experience. Remote exams risk to exacerbate the bureaucratic aspects of the examination process at the expense of the interaction between lecturers and students that should occur in the context of an exam. The perception of the exam as a pure bureaucratic obligation emphasizes students' attention toward the exam outcome, thus contributing to obscure the more important motivations related to the learning aims. This consideration, on the one hand, concerns aspects that are less tangible than the exam outcome, such as the relationship between evaluation and motivation, and arises more general issues about the role of examination. On the other hand, it encourages to think about exams in a different way and to question if, despite being taken face-to-face or online, they already are a mere bureaucratic procedure that creates a real distance between lecturers and students.

REFERENCES

- Agresti, A. (2003). *Categorical data analysis*. John Wiley & Sons, New York.
- Agresti, A. (2018). *An introduction to categorical data analysis*. John Wiley & Sons, New York.
- Bologna Declaration. (1999). *The European higher education area*. Joint declaration of the European ministers of education convened in Bologna at the 19th of June 1999, Bologna.
- Davies, L. and Bentrovato, D. (2011). *Understanding education's role in fragility: Synthesis of four situational analyses of education and fragility: Afghanistan, Bosnia and Herzegovina, Cambodia, Liberia*. International institute for educational planning. UNESCO.
- Deleeuw, J., Goldstein, H. and Meijer, E. (2007). *Handbook of Multilevel Analysis*, Springer, New York.
- Edgington, E. and Onghena, P. (2007). *Randomization tests*. CRC Press.
- George, M.L. (2020). Effective teaching and examination strategies for undergraduate learning during COVID-19 school restrictions. *Journal of Educational Technology Systems*, 49(1): 23-48.
- Gonzalez, T., De la Rubia, M.A., Hincz, K.P., Comas-Lopez, M., Subirats, L., Fort, S. and Sacha, G.M. (2020). Influence of COVID-19 confinement on students' performance in higher education. *PloS one*, 15(10), e0239490.
- Good, P. (2013). *Permutation tests: a practical guide to resampling methods for testing hypotheses*. Springer Science & Business Media, Berlin-Heidelberg.

- Means, B., Bakia, M. and Murphy, R. (2014). *Learning online: What research tells us about whether, when and how*. Routledge, New York.
- Merrill, M.D., Van Merriënboer, J.J. and Driscoll, M.P. (2008). *Handbook of research on educational communications and technology* (Vol. 3). J.M. Spector (Ed.). New York: Lawrence Erlbaum Associates.
- OECD. (2020). *Learning remotely when schools close: How well are students and schools prepared? Insights from PISA*. OECD Policy Responses to Coronavirus (Covid-19).
- Porta M. (2008). *A Dictionary of Epidemiology*. Oxford University Press, New York.
- Rabe-Hesketh, S. and Skrondal, A. (2008). *Multilevel and longitudinal modeling using Stata*. STATA press.
- Robinson, C.C. and Hullinger, H. (2008). New benchmarks in higher education: Student engagement in online learning, *Journal of Education for Business*, 84(2): 101-109.
- Simonds, T.A. and Brock, B.L. (2014). Relationship between age, experience, and student preference for types of learning activities in online courses. *Journal of Educators Online*, 11(1): 1-19
- Snijders, T.A.B. and Bosker, R. J. (1999). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Sage, London.
- Stack, S. (2015). Learning Outcomes in an online vs traditional course. *International Journal for the Scholarship of Teaching and Learning*, 9(1): 1-20.

APPENDIX A

Number of studies programmes at the University of Trento, listed by cycle programme and department.

	First cycle	Second cycle	Single Cycle
Department of Economics and Management	3	7	-
Department of Civil, Environmental and Mechanical engineering	2	4	1
Department of Law	1	-	1
Department of Information engineering and Computer science	3	3	-
Department of Humanities	4	5	-
Department of Industrial engineering	1	1	-
Department of Psychology and Cognitive science	2	3	-
Department of Sociology and Social research	3	3	-
Department of Mathematics	1	2	-
Department of Physics	1	1	-
Department of Cellular, Computational and Integrative Biology	1	2	-
Center of Agriculture Food Environment	1	-	-
School of International Studies	-	2	-
TOTAL	23	33	2