Frontiers in Education

TYPE Original Research
PUBLISHED 01 June 2023
DOI 10.3389/feduc.2023.1160430

OPEN ACCESS

EDITED BY
Aleksander Aristovnik,
University of Ljubljana, Slovenia

REVIEWED BY
Omar Mahasneh,
Al-Balqa Applied University, Jordan
Sohaib Ahmed,
Bahria University, Pakistan
Bibhya Sharma,
University of the South Pacific, Fiji

*CORRESPONDENCE
Guosheng Yin
✉ gyin@hku.hk

RECEIVED 07 February 2023
ACCEPTED 15 May 2023
PUBLISHED 01 June 2023

CITATION
Cheung YYH, Lam KF, Zhang H, Kwan CW, Wat KP, Zhang Z, Zhu K, Chung YK and Yin G (2023) A randomized controlled experiment for comparing face-to-face and online teaching during COVID-19 pandemic. *Front. Educ.* 8:1160430.
doi: 10.3389/feduc.2023.1160430

A randomized controlled experiment for comparing face-to-face and online teaching during COVID-19 pandemic

Yue Yat Harrison Cheung[1], Kwok Fai Lam[1,2], Hengtao Zhang[1], Chi Wai Kwan[1], Kam Pui Wat[1], Zhiqiang Zhang[1], Ke Zhu[1], Yuk Ka Chung[1,3] and Guosheng Yin[1,4]*

[1]Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong, China, [2]Centre for Quantitative Medicine, Duke-NUS Medical School, Singapore, Singapore, [3]Department of Mathematics, The University of Manchester, Manchester, United Kingdom, [4]Department of Mathematics, Imperial College London, London, United Kingdom

Randomized controlled experiments have shown that face-to-face teaching is more effective in delivering various learning outcomes than asynchronous online teaching. Unlike the asynchronous online teaching mode, the synchronous online mode has a live instruction component and is more comparable to the face-to-face mode. A small-sized randomized controlled experiment involving 50 students showed that there was no significant difference in student ratings on the effectiveness between the face-to-face and synchronous online teaching modes. Prior to the current study, no medium-or large-sized randomized controlled experiment had been conducted for comparing the two modes. The current study aims to fill in the gap by comparing the effectiveness of face-to-face (i.e., intervention) and synchronous online (i.e., control) teaching through a randomized controlled experiment involving 725 students from seven statistics courses offered by the Department of Statistics and Actuarial Science at the University of Hong Kong. Results show that the difference in learning outcomes between the two modes is not statistically significant. The class size is an effect modifier that students assigned to the face-to-face mode have significantly higher final weighted and final exam scores if they have face-to-face lessons with 25 students or fewer. The Pass/Fail grading option has a significantly negative effect on course performance.

KEYWORDS

e-learning, face-to-face learning, online teaching, randomized controlled experiment, teaching effectiveness

Introduction

Face-to-face teaching has been an educational norm for thousands of years. Through face-to-face communications, students and teachers get acquainted with each other, and teaching and learning activities are not only for the sake of knowledge transmission but also involve mutual influence of life attitudes and personalities. Before entering society, students spend more than 10 years at schools and colleges where they learn how to communicate with their peers and seniors, obey the social norms, and acquire theoretical knowledge and practical skills for entering their future professions. Therefore, face-to-face teaching used to be considered indispensable by teachers and students (Smith et al., 2009; Sithole et al., 2019; Morreale et al., 2021). However, things have changed drastically since the outbreak of COVID-19 pandemic. Similar to many industries, educational activities have been interrupted and many schools and universities have been temporarily shut down before eventually moving to online platforms (Abbasi et al., 2020; Aboagye et al., 2020; Radha et al., 2020; Mathivanan et al., 2021). Millions of university-level courses are being delivered online across the world, and this trend has lasted for a long period making us wonder whether e-learning will eventually overtake face-to-face learning and become a new educational norm in the post-COVID-19 era (Saeed Al-Maroof et al., 2020; Hermawan, 2021; Pham and Da Vo, 2021; Shofwan et al., 2021; Tang et al., 2021).

In the next section, we provide a brief overview of the motivation behind this research. We then review past literature on the comparison between face-to-face and online teaching with learning theories and details of the learning outcome framework. In the subsequent sections, we discuss the methodology, results and generalizability of this study.

Motivation

Before the pandemic, courses from higher education in many parts of the world already had some online elements in the sense that many universities used learning management systems such as Moodle and Canvas, where students could download materials, watch teaching videos, ask questions and receive instructions from the course forum (Anand and Eswaran, 2018; Grossi et al., 2018). Therefore, a natural question is whether the face-to-face element still has some added value over a purely online teaching mode. In this paper, the added value is confined to that related to course performance, which can be measured by assignment, test and final exam scores.

Literature review

Teaching and Learning with information communication technology

Past studies showed that e-learning had suffered from technical deficiencies such as the instability of spontaneous communication and long loading time for educational videos (Nwabufo et al., 2013; Aminu and Rahaman, 2014; Al-Azawei et al., 2016). Many of these issues have been resolved by the advancement of Information Communication Technology (ICT). Sharma et al. (2018) reported that smart learning tools have been adopted by the University of the South Pacific, leading to a healthy smart learning ecosystem. Reddy et al. (2021) pointed out that computer competency and computer self-efficacy were the two major factors determining the acceptance of smart learning technology among secondary school students and university newcomers. This echoed the key finding of their previous paper (Reddy et al., 2020) that freshmen had a high digital literacy level in general. However, Khalil et al. (2020) stressed the fact that online students might be unable to concentrate properly without eye contact with the teachers.

Learning theories

There are competing learning theories for face-to-face versus online teaching. While the social learning theory suggests face-to-face teaching provides opportunities for direct social interactions and collaborations that can facilitate teaching and learning (Bandura and Walters, 1977), the cognitive load theory (CLT) suggests online teaching can be more effective in reducing the cognitive loads of students by allowing them to study in their own pace and time (Hadie et al., 2021). Finally, constructivism suggests learning is a self-constructive process and therefore, both face-to-face and online teaching can help students construct knowledge from their own learning experience (Kleinke and Lin, 2020; Fatimah et al., 2022).

Observational studies

Observational studies, both qualitative and quantitative in nature, were conducted to compare the effectiveness of face-to-face, online, and blended teaching with mixed results (Larson and Sung, 2009; Lu and Lemonde, 2013; Broadbent, 2017; Ebner and Gegenfurtner, 2019; Herodotou et al., 2020; Littenberg-Tobias and Reich, 2020; Randazzo et al., 2021). Mahasneh et al. (2022) conducted a survey on a sample of 3,584 students and found that the most important positive effect of online teaching was the improvement in students' ability to use smart devices for educational purposes. Tawarah et al. (2022) pointed out that the availability of facilities and students' self-motivation for study were the most important determinants for achieving online teaching objectives.

Randomized controlled experiments

Small to medium-sized randomized controlled experiments were also conducted to compare face-to-face, blended, and asynchronous online teaching (Alpert et al., 2016; Reavley et al., 2018; Crawford et al., 2021; Jiang et al., 2021). In those studies, blended and face-to-face teaching modes were shown to be more effective in delivering learning outcomes than the asynchronous online teaching mode. Alnabelsi et al. (2015) conducted a randomized controlled experiment with 50 students for comparing the effectiveness of face-to-face and synchronous online teaching, and there was no significant difference in student ratings between the two teaching modes for the usefulness of the lecture. So far, no medium-or large-sized randomized controlled experiment has been conducted for comparing face-to-face and synchronous online teaching. The goal of this paper is to fill in this gap. To the best of our knowledge, our study is by far the largest randomized controlled experiment for comparing face-to-face and online teaching.

Learning outcome frameworks

Influenced by Bloom's taxonomy of educational objectives (Krathwohl, 2002), two popular frameworks have been adopted by the aforementioned studies, that is, the grade-and survey-based learning outcome frameworks. While the grade-based framework uses assessment scores such as assignment and exam results to assess the

10.3389/feduc.2023.1160430

fulfillment of learning outcomes, the survey-based framework mainly relies on surveys or other forms of student feedback for the evaluation of student perceived learning outcomes and satisfaction. The current study adopts the former learning outcome framework, and the research question is whether face-to-face teaching can enable students to achieve significantly higher assessment scores than synchronous online teaching.

Methodology

Experimental design and data collection

The experiment started on September 1, 2020 and ended on December 23, 2020. The intervention started from the 4th week of the semester, i.e., on September 21, 2020. Consent forms were sent to students before the experiment, and only the students who signed the consent forms were included in the sample. The sample consisted of 725 students from seven statistics courses offered by the University of Hong Kong in the first semester of 2020–2021, among which 560 and 165 are undergraduate and postgraduate students, respectively. For the two large courses with enrollment over 140 (i.e., Courses A and B), students were randomly assigned to either the face-to-face (i.e., intervention) or synchronous online (i.e., control) teaching mode in ratios of 1:2 and 1:3, respectively, whereas students in other smaller courses were randomly assigned to the two modes in a ratio of 1:1. The difference in the assignment ratio was to comply with the COVID-19 measure that the number of students in a classroom could not exceed one third of its maximum capacity. A stratified randomization was adopted for this study with course code as the stratification factor.

Lecture and tutorial materials were posted on the learning management system of Moodle and were accessible to both face-to-face and synchronous online mode students. In each week, 2–3 h of lectures and 1 h of tutorial were conducted, during which the face-to-face mode students received face-to-face instructions, whereas the synchronous online mode students received identical instructions *via* Zoom simultaneously. Hence, the only difference between the face-to-face and synchronous online teaching modes is that the face-to-face mode students received face-to-face instructions in lectures and tutorials, whereas the synchronous online mode students received the same instructions *via* Zoom. Both groups of students had access to the recorded teaching videos.

The first 2 weeks of the semester were the add/drop period during which students in the seven courses were informed to participate in the study and their informed consents were collected. In the third week of the semester, an online pretest was given to all students in the experiment, and it served as a baseline variable indicating students' ability and understanding of course prerequisite knowledge. Owing to the COVID-19 pandemic, students were allowed to choose between the Pass/Fail and Letter grading options for each course taken and would receive only a Pass/Fail grade on the transcript for the course if the Letter grading option was opted out. There were two class tests (excluding the pretest) and 3–5 take-home assignments, and an online final exam for each course. The five outcome variables (i.e., the final weighted score, final exam, coursework, test and assignment scores) along with other baseline variables such as gender, grading option, pretest score, course code, course level (undergraduate vs. postgraduate) and technicality (technical vs. non-technical course content) were collected for the purpose of this study. While the final weighted score was a weighted average of the final exam and coursework scores, the coursework score was a weighted average of the assignment and test scores. The final exam score was still collected even if a student opted for the Pass/Fail grading option.

Statistical analyses were conducted in the following steps. First, we compared the baseline characteristics between the intervention and control groups in the section of descriptive statistics. Second, we tested the null hypothesis that there is no difference in the effectiveness between face-to-face and synchronous online teaching in delivering the assessment outcomes in the initial part of the results section. A forest plot was drawn for the subgroup analysis on the treatment effect of the delivery mode. Third, we investigated the interaction effect between the delivery mode and class size and tested the null hypothesis that there is no interaction between the delivery mode and class size in the latter part of the results section. Finally, we compared students' final exam and pre-exam performances so as to provide some additional suggestions for improving teaching quality in the post-pandemic era.

Descriptive statistics

The statistical software of R (version 4.1.0) and JMP (version Pro 16.0.0) were used to conduct the analyses in this paper, which include descriptive statistics in the current section, and linear regressions and a forest plot in the next section.

In Table 1, we compare the baseline characteristics between the intervention and control groups on the pretest score, percentage of students opted for Pass/Fail grading, gender, course level and technicality, and the *ps*-value from two-sample t-tests (i.e., the 2nd row) and *Z*-tests for proportions (i.e., from the 3rd to 6th rows) are presented in the final column. All hypothesis tests are two-sided, and the 95% confidence intervals (CI) for the mean pretest scores are given on the second row. While the first three are student-endogenous variables that have been well balanced by stratified

TABLE 1 Comparison of baseline characteristics between intervention (face-to-face) and control (online) groups.

Baseline variable	Face-to-face (n=282)	Online (n=443)	p-value
Pretest Score 95% CI	66.62 [65.26, 67.98]	66.52 [65.43, 67.61]	0.955
Opted for Pass/Fail Grading	107/282 (37.9%)	144/443 (32.5%)	0.134
Gender (male %)	163/282 (57.8%)	270/443 (61.0%)	0.400
Course Level (UG Student %)	199/282 (70.6%)	361/443 (81.5%)	0.001
Technicality (Technical %)	173/282 (61.4%)	171/443 (38.6%)	0.000

Pretest score is a baseline indicator for students' understanding of course prerequisite knowledge. Students can choose to have either Letter Grading or Pass/Fail Grading. There are two course levels, the undergraduate (UG) and postgraduate (PG). Two large non-technical undergraduate courses (i.e., Courses A and B) have lower face-to-face to online assignment ratios due to the COVID-19 social distancing measures, which leads to the baseline imbalance in the last two rows. This is addressed by controlling course code variables in the multiple regression so that the course-specific effects of technicality and course level are captured by the coefficients of the course code variables.

randomization, the last two are course specific student-exogenous variables (Guney, 2009). Technical courses are defined to be those involving more theories, mathematical derivations, and proofs. Owing to the COVID-19 social distancing measures, the two large non-technical courses (i.e., courses A and B) had lower face-to-face to online ratios (1:2 and 1:3) than other smaller technical courses (i.e., courses C to G). Since randomization was performed within each course, the student-exogenous variables (i.e., technicality and course level) were not intended to be balanced by randomization. Therefore, the face-to-face group had a significantly lower proportion of undergraduate students but a higher proportion of students taking technical courses. Furthermore, it is a well-known fact that courses could have different mean scores due to different technical levels and grading standards of teachers. Hence, we address the course specific differences by controlling for the course code in the multiple regression. That is, the course-specific effects of technicality, course level and teacher's standard are captured by the coefficients of the course code variables.

More information on the relevant courses is given in Table 2, where Courses A and B are non-technical courses, and all other courses are relatively more technical. Due to large class sizes and social distancing requirements during pandemic, these two undergraduate courses used assignment ratios of 1:2 and 1:3, respectively, for face-to-face versus synchronous online teaching modes. Gender and course level distributions of the participants are presented in Figure 1.

Results

Multiple linear regression models are fitted to the intention-to-treat (ITT) data and the ITT effects of the delivery mode on learning outcomes are estimated. We consider the final weighted score, final exam, test average, coursework and assignment average scores separately as outcome variables and examined their associations with the delivery mode (Face-to-face vs. Online), grading option (Pass/Fail vs. Letter grade), gender and course code with course G as the reference group. Variance inflation factor (VIF) has been used to measure the degree of multicollinearity in the multiple regression models. None of the VIFs is greater than 3, showing the degree of multicollinearity is low. From Table 3, we have the following key findings:

- The effect of the delivery mode on the five assessment measures was not statistically significant. We cannot reject the null hypothesis that there is no difference in effectiveness between the face-to-face and online teaching modes.
- Students with higher pretest scores performed significantly better in all five assessment measures.
- Students who opted for a Pass/Fail grade had significantly lower final weighted, final exam, coursework, test average and assignment average scores than those who opted for a Letter grade.
- Male students had significantly lower assignment scores than female students.

We further investigate the subgroup treatment effect of the delivery mode with course code as a stratification variable. The forest plot in Figure 2 shows that, the subgroup treatment effect for each course was not statistically significant except that Course A students performed better in assignments if they were assigned to the face-to-face mode.

Inspired by the argument that eye contact with the teacher can help student's concentration (Khalil et al., 2020), we investigate whether the class size has been a significant modifier for the effect of delivery mode. Understandably, most students would not have eye contact with the teacher in a large class, and face-to-face teaching might not help students' concentration. Under the potential outcome framework (Rubin, 1974), we assume that

$$E\{Y(z)|V,\mathbf{x}\} = \beta_0 + \beta_1 z + \beta_2 I(V \le s) \times z + \beta_3^T \mathbf{x},$$

where $Y(z)$ is the potential learning outcome given the delivery mode with $z=1$ referring to the face-to-face mode, and $z=0$ the online mode, $I(\cdot)$ is an indicator function and is equal to one when the class size V (i.e., the face-to-face # in Table 2) is less than or equal to a certain threshold s, and zero otherwise, and $\mathbf{x}$ is a vector of other baseline covariates such as the pretest score, grading option, gender and course code. The treatment effect of the delivery mode on the learning outcome with adjustment of the face-to-face class size is given by

TABLE 2 Course information with the numbers of students in the face-to-face and online teaching modes.

Course code	Course name	Course level	Technicality	Face-to-face #	Online #
A	Statistics: Ideas and Concepts	Undergraduate	Non-technical	50	97
B	Introductory Statistics	Undergraduate	Non-technical	59	175
C	Probability and Statistics 1	Undergraduate	Technical	61	52
D	Probability and Statistics 2	Undergraduate	Technical	23	32
E	Computational Statistics	Postgraduate	Technical	20	17
F	Statistical Inference for Data Science	Postgraduate	Technical	36	38
G	Advanced Statistical Modeling	Postgraduate	Technical	33	32

The face-to-face # refers to the class size of face-to-face lectures and tutorials.

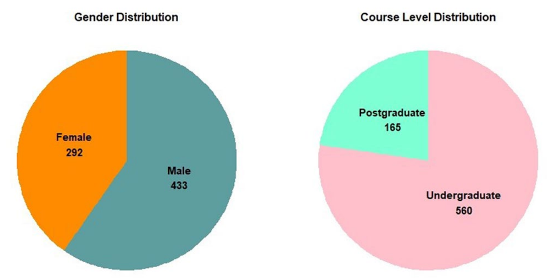

FIGURE 1
Gender and course level distributions of the participants.

$$E\{Y(1)|V,\mathbf{x}\} - E\{Y(0)|V,\mathbf{x}\} = \beta_1 + \beta_2 I(V \le s).$$

We investigate the interaction effect between the delivery mode and class size with different thresholds. As shown in Figure 3, there are only two outcome variables (i.e., the final weighted and final exam scores) with statistically significant interaction effects when the class size threshold $s = 25$. For these two outcome variables, the estimated interaction effect first increases and then decreases from the class size threshold $s = 25$ onward, while the corresponding p-value reaches the minimum at $s = 25$. For the other outcomes, similar patterns can be observed, although they never reach the 5% significance level. For $s = 20$, only one course (i.e., Course E) falls into this category. Note

TABLE 3 Multiple regression analyses of five assessment measures with the delivery mode, pretest score, grading option, gender, and course code as covariates.

Predictor	Outcome				
	Final weighted	Final exam	Coursework	Test	Assignment
Mode [Face-to-face]	−0.51	−0.33	0.08	−0.4	1.48
95% CI	[−2.37, 1.35]	[−2.57, 1.92]	[−1.77, 1.93]	[−2.45, 1.65]	[−0.68, 3.65]
p-value	0.592	0.776	0.934	0.701	0.18
Pretest	0.4	0.4	0.44	0.52	0.26
95% CI	[0.35, 0.45]	[0.34, 0.46]	[0.39, 0.49]	[0.46, 0.58]	[0.20, 0.32]
p-value	0	0	0	0	0
Pass/Fail	−9.45	−10.87	−6.36	−7.51	−4.32
95% CI	[−11.46, −7.44]	[−13.29, −8.44]	[−8.36, −4.37]	[−9.73, −5.30]	[−6.66, −1.97]
p-value	0	0	0	0	0
Male	−0.64	−0.73	−0.92	0.01	−2.29
95% CI	[−2.45, 1.17]	[−2.91, 1.44]	[−2.71, 0.88]	[−1.97, 1.99]	[−4.39, −0.19]
p-value	0.489	0.509	0.316	0.995	0.033
Course A	−5.04	−7.15	−6.85	−14.58	−1.14
95% CI	[−7.12, −2.96]	[−9.65, −4.65]	[−8.90, −4.79]	[−16.86, −12.29]	[−3.55, 1.26]
p-value	0	0	0	0	0.351
Course B	0.85	6.09	0.41	−2.38	1.1
95% CI	[−0.95, 2.64]	[3.93, 8.25]	[−1.38, 2.19]	[−4.35, −0.41]	[−0.98, 3.19]
p-value	0.356	0	0.655	0.018	0.299
Course C	−1.28	−13.23	5.55	5.79	1.99
95% CI	[−3.52, 0.97]	[−15.93, −10.53]	[3.32, 7.78]	[3.32, 8.25]	[−0.61, 4.59]
p-value	0.265	0	0	0	0.135
Course D	−6.19	−0.01	−3.81	−7.03	−6.64
95% CI	[−9.25, −3.12]	[−3.70, 3.67]	[−6.86, −0.76]	[−10.39, −3.67]	[−10.19, −3.09]
p-value	0	0.994	0.015	0	0
Course E	−3.68	−12.57	−2.07	12.4	−11.93
95% CI	[−7.33, −0.04]	[−16.96, −8.18]	[−5.70, 1.57]	[8.41, 16.40]	[−16.16, −7.71]
p-value	0.048	0	0.266	0	0
Course F	7.08	14.02	3.95	0.64	8.18
95% CI	[4.56, 9.61]	[10.98, 17.06]	[1.43, 6.47]	[−2.13, 3.42]	[5.24, 11.13]
p-value	0	0	0.002	0.651	0

Course G is used as the reference group. The final weighted score is a weighted average of the final exam and coursework scores. The coursework score is a weighted average of the assignment and test scores.

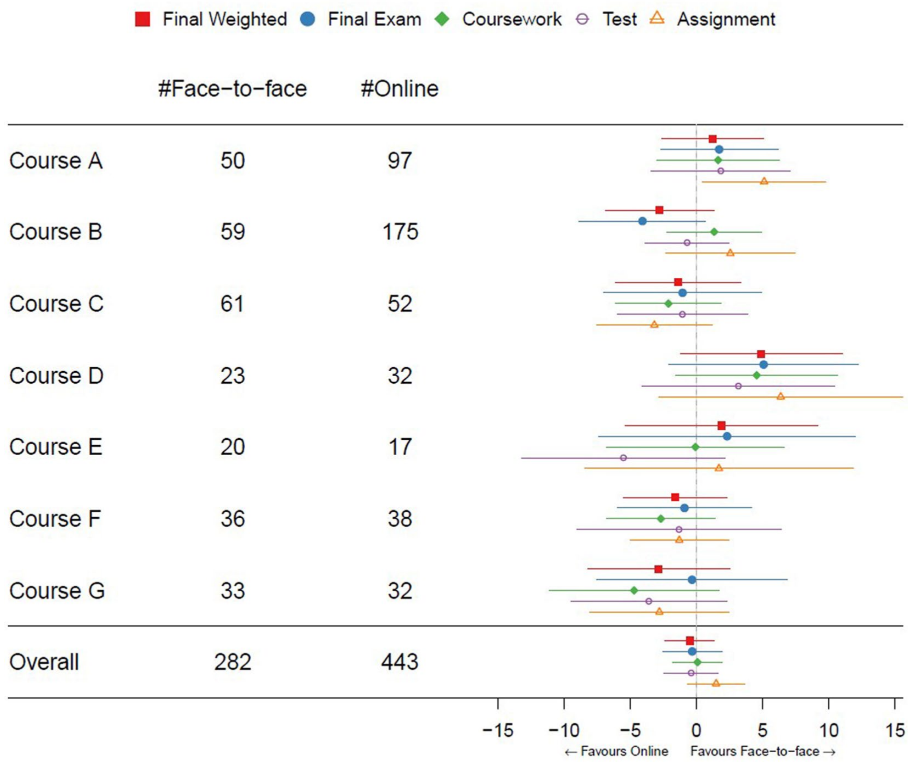

FIGURE 2
Subgroup treatment effect by the forest plot with course code as the stratification variable.

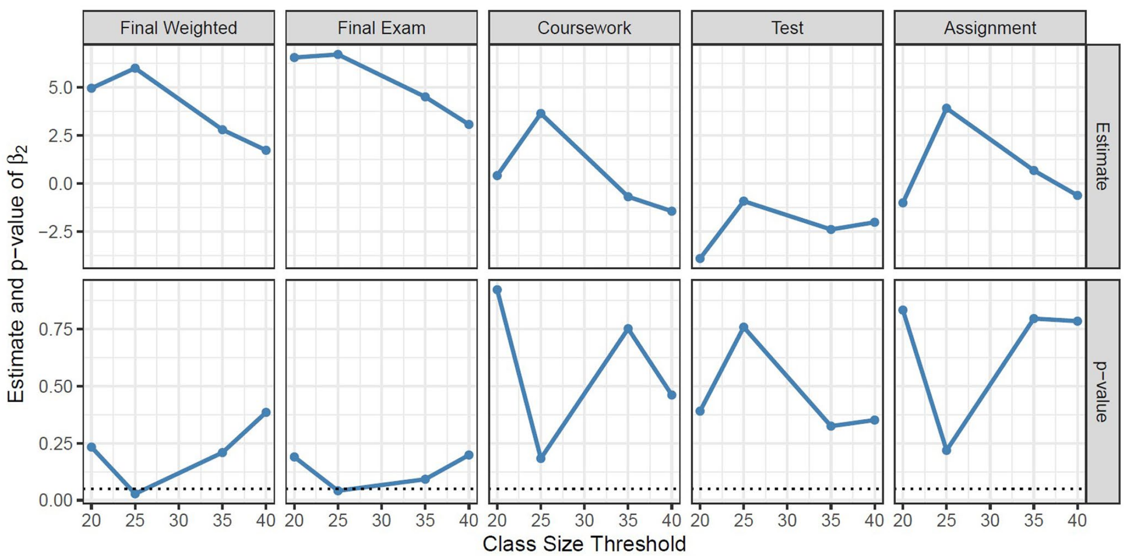

FIGURE 3
Estimates for the interaction effect between the delivery mode and class size with different class size thresholds.

that β_2 for this category is not statistically significant and the trend is not shown due to a small sample size in this threshold category. Based on the findings in Figure 3, we conclude that $s = 25$ would be the optimal class size threshold.

Table 4 shows that, students who were assigned to the face-to-face mode had significantly higher final weighted and final exam scores if they had face-to-face lectures and tutorials in small classes with 25 students or fewer. The corresponding estimators β_2's are positive (5.99 and 6.70) and statistically significant (*ps*-value: 0.029 and 0.042). There is strong evidence suggesting that the class size interacts with the delivery mode and that face-to-face teaching is more effective than synchronized online teaching if the class size is small.

TABLE 4 Estimates for the effect of the delivery mode on learning outcomes with the class size as a modifier.

Estimate	Final weighted	Final exam	Coursework	Test	Assignment
$\hat{\beta}_1$ 95% CI	−1.33 [−3.33, 0.67]	−1.24 [−3.65, 1.16]	−0.41 [−2.40, 1.57]	−0.27 [−2.48, 1.93]	0.95 [−1.38, 3.27]
p-value	0.193	0.311	0.683	0.808	0.426
$\hat{\beta}_2$ 95% CI	5.99 [0.64, 11.34]	6.70 [0.26, 13.14]	3.63 [−1.71, 8.97]	−0.93 [−6.82, 4.97]	3.90 [−2.31, 10.12]
p-value	0.029	0.042	0.183	0.758	0.219

Finally, Figure 4 compares the final exam and pre-exam (i.e., coursework including tests and assignments) performances of students with the delivery mode as a stratification variable. The generalized likelihood ratio test shows little difference in the slopes of the two regression lines (Face-to-face vs. Online), which implies the association between the final exam and pre-exam performances was the same for the two teaching modes (p-value = 0.575). Both regression lines have positive slopes and are below the 45-degree dashed line, implying that coursework performance was positively correlated with final exam performance, and that students in both teaching modes performed better in coursework than in the final exam (because students were allowed to discuss and refer to textbooks and notes for coursework). As a form of continuous assessment, a more rigorous marking scheme for coursework can encourage students to work hard throughout the entire semester.

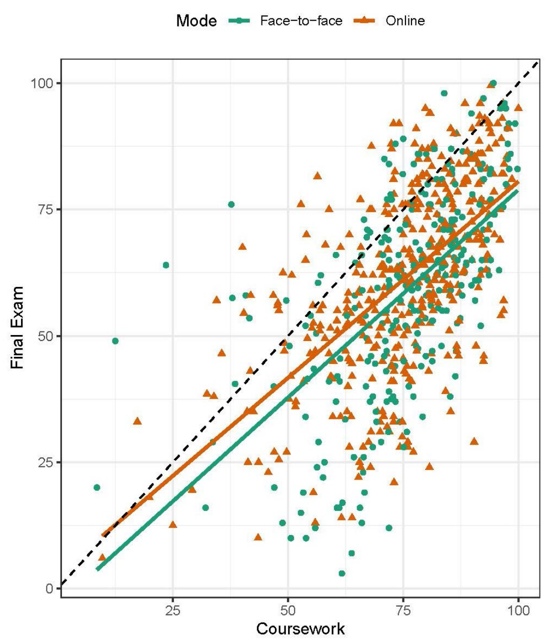

FIGURE 4
Comparison between the final exam and pre-exam performances of students with the delivery mode as the stratification variable.

Discussion

Since the COVID-19 outbreak, there has been a distinctive rise of e-learning, whereby teaching is undertaken remotely on digital platforms. Will e-learning eventually overtake face-to-face teaching and become a new educational norm in the post-COVID-19 era? Does face-to-face teaching still have some added value over a purely online teaching mode? How to improve teaching quality in the post-COVID-19 era? To provide some good insights into these questions, if not answering them, we conducted a randomized controlled experiment comparing the effectiveness of face-to-face and synchronous online teaching.

There has been insufficient evidence to suggest the face-to-face element can enable students to perform better in statistics courses at the university level. This result agrees with that of a randomized controlled experiment conducted by Alnabelsi et al. (2015). Note that our sample size is much larger, and we used assessment scores as means of comparison, which may be more objective than the use of student ratings in the previous study. In our sample, three out of seven courses had face-to-face class sizes of more than 50. Since it was not unusual for large classes to be noisy, we suspect that students in large classes could have distracted each other. Moreover, students in small face-to-face classes could have more interactions with the teacher, which was an advantage they had over students in the online teaching group. The class size was an effect modifier that students assigned to the face-to-face mode had significantly higher final weighted and exam scores (on average), if the face-to-face lessons were taught in small classes with 25 students or fewer. Finally, the Pass/Fail grading option had a significantly negative effect on course performance. Once chosen the Pass/Fail grading option, students stopped working hard because assessments would not affect their GPAs given they managed to pass the course.

A limitation of our study is the fact that all students who participated in this experiment were taking statistics courses, and our results might not be applicable to the teaching of other subjects. This is especially true for those subjects that are different from statistics in nature, such as history, literature, music, and visual arts. However, our study can provide some insights into how teaching could be improved in the post-COVID-19 era. Since the face-to-face element cannot improve students' assessment scores, we provide the following suggestions.

- Teaching videos can still be recorded in the post-COVID-19 era so that students can re-watch the technical parts of the course and gain a better understanding. These videos are particularly beneficial to the students who are absent from the classes due to various reasons such as medical issues, professional leave or conference attendance.
- Online teaching provides flexibilities to part-time students who have rigid working hours and cannot freely travel to campus for in-person learning.
- Pre-exam assessments (i.e., tests and assignments) can account for higher weights with more stringent marking standards so that students are encouraged to work hard throughout the entire

semester, rather than making an imbalanced, last-minute effort right before the final exam.

- Face-to-face teaching should be offered in small classes with more student engagements for its value to be fully realized. Small classes can also allow students to be more acquainted with each other and build up positive peer learning relationships.
- Face-to-face teaching, if there is any, should involve more interactions between teachers and students and focus more on the intangible learning outcomes such as motivation for study, encouragement of creativity and the quest for better humanity.
- In the era of big data and artificial intelligence, the value of face-to-face teaching lies less on knowledge transmission but more on the intangible learning outcomes that can enrich humanity and help distinguish humans from machines.

Conclusion

In this study, we conduct a randomized controlled experiment to compare the effectiveness of face-to-face and synchronous online teaching. There has been insufficient evidence to suggest that the learning outcomes of the two modes are different. The class size is a significant effect modifier that students assigned to the face-to-face mode have significantly higher final weighted and exam scores if they have face-to-face lessons with 25 students or fewer.

Data availability statement

All data and code related to this paper have been made publicly available at the Open Science Framework (OSF) and can be accessed via: https://doi.org/10.17605/OSF.IO/3Q7PZ.

Ethics statement

The studies involving human participants were reviewed and approved by Institutional Review Board of the University of Hong Kong. The participants provided their written informed consent to participate in this study.

Author contributions

KL and YC have jointly written the first draft and have equal contributions to the paper. GY and HZ validated the results. GY and KL revised the drafts. HZ produced the 3 figures in the paper. All other authors have offered constructive comments during the revision process and participated in the data collection and design of the study. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abbasi, S., Ayoob, T., Malik, A., and Memon, S. I. (2020). Perceptions of students regarding e-learning during Covid-19 at a private medical college. *Pakistan J. Med. Sci.* 36, S57–S61. doi: 10.12669/pjms.36.COVID19-S4.2766

Aboagye, E., Yawson, J. A., and Appiah, K. N. (2020). COVID-19 and e-learning: the challenges of students in tertiary institutions. *Soc. Educ. Res.* 2, 1–8. doi: 10.37256/ser.212021422

Al-Azawei, A., Parslow, P., and Lundqvist, K. (2016). Barriers and opportunities of e-learning implementation in Iraq: a case of public universities. *Int. Rev. Res. Open Distrib. Learn.* 17, 126–146. doi: 10.19173/irrodl.v17i5.2501

Alnabelsi, T., Al-Hussaini, A., and Owens, D. (2015). Comparison of traditional face-to-face teaching with synchronous e-learning in otolaryngology emergencies teaching to medical undergraduates: a randomised controlled trial. *Eur. Arch. Otorhinolaryngol.* 272, 759–763. doi: 10.1007/s00405-014-3326-6

Alpert, W. T., Couch, K. A., and Harmon, O. R. (2016). A randomized assessment of online learning. *Am. Econ. Rev.* 106, 378–382. doi: 10.1257/aer.p20161057

Aminu, H., and Rahaman, S. (2014). Barriers thrusting e-learning to the backseat: Nigeria a case study. In 2014 IEEE Canada International Humanitarian Technology Conference-(IHTC) (pp. 1–4). IEEE.

Anand, A., and Eswaran, S. (2018). A survey of open source learning management systems. *Ann. Comp. Sci. Ser.* 16, 185–188.

Bandura, A., and Walters, R. H. (1977). *Social learning theory* (*1*). Prentice Hall: Englewood cliffs.

Broadbent, J. (2017). Comparing online and blended Learner's self-regulated learning strategies and academic performance. *Internet High. Educ.* 33, 24–32. doi: 10.1016/j.iheduc.2017.01.004

Crawford, A., Varghese, C., Hsu, H. Y., Zucker, T., Landry, S., Assel, M., et al. (2021). A comparative analysis of instructional coaching approaches: face-to-face versus remote coaching in preschool classrooms. *J. Educ. Psychol.* 113, 1609–1627. doi: 10.1037/edu0000691

Ebner, C., and Gegenfurtner, A. (2019). Learning and satisfaction in webinar, online, and face-to-face instruction: a Meta-analysis. *Front. Educ.* 4:92. doi: 10.3389/feduc.2019.00092

Fatimah, S., Rosidin, D. N., and Hidayat, A. (2022). Student-based learning in the perspective of constructivism theory and Maieutics method. *Int. J. Soc. Sci. Hum. Res.* 05, 1632–1637. doi: 10.47191/ijsshr/v5-i5-10

Grossi, M. G. R., Elias, M. C. A. S., Chamon, C. M., and Leal, D. C. C. C. (2018). The educational potentialities of the virtual learning environments Moodle and canvas: a comparative study. *Int. J. Inform. Educ. Technol.* 8, 514–519. doi: 10.18178/ijiet.2018.8.7.1091

Guney, Y. (2009). Exogenous and endogenous factors influencing students' performance in undergraduate accounting modules. *Acc. Educ.* 18, 51–73. doi: 10.1080/09639280701740142

Hadie, S. N. H., Tan, V. P. S., Omar, N., Nik Mohd Alwi, N. A., Lim, H. L., and Ku Marsilla, K. I. (2021). COVID-19 disruptions in health professional education: use of cognitive load theory on students' comprehension, cognitive load, engagement, and motivation. *Front. Med.* 8:739238. doi: 10.3389/fmed.2021.739238

Hermawan, D. (2021). The rise of e-learning in COVID-19 pandemic in private university: challenges and opportunities. IJORER: international journal of recent. *Educ. Res.* 2, 86–95. doi: 10.46245/ijorer.v2i1.77

Herodotou, C., Muirhead, D. K., Aristeidou, M., Hole, M. J., Kelley, S., Scanlon, E., et al. (2020). Blended and online learning: a comparative study of virtual microscopy in higher education. *Interact. Learn. Environ.* 28, 713–728. doi: 10.1080/10494820.2018.1552874

Jiang, D., Kalyuga, S., and Sweller, J. (2021). Comparing face-to-face and computer-mediated collaboration when teaching EFL writing skills. *Educ. Psychol.* 41, 5–24. doi: 10.1080/01443410.2020.1785399

Khalil, R., Mansour, A. E., Fadda, W. A., Almisnid, K., Aldamegh, M., Al-Nafeesah, A., et al. (2020). The sudden transition to synchronized online learning during the COVID-19 pandemic in Saudi Arabia: a qualitative study exploring medical students' perspectives. *BMC Med. Educ.* 20:285. doi: 10.1186/s12909-020-02208-z

Kleinke, S., and Lin, Y. (2020). Application of adult learning theory to STEM education in online learning environment. In 2020 IEEE Frontiers in education conference (FIE) IEEE.

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: an overview. *Theory Pract.* 41, 212–218. doi: 10.1207/s15430421tip4104_2

Larson, D. K., and Sung, C. H. (2009). Comparing student performance: online versus blended versus face-to-face. *JALN* 13, 31–42. doi: 10.24059/olj.v13i1.1675

Littenberg-Tobias, J., and Reich, J. (2020). Evaluating access, quality, and equity in online learning: a case study of a MOOC-based blended professional degree program. *Internet High. Educ.* 47:100759. doi: 10.1016/j.iheduc.2020.100759

Lu, F., and Lemonde, M. (2013). A comparison of online versus face-to-face teaching delivery in statistics instruction for undergraduate health science students. *Adv. Health Sci. Educ.* 18, 963–973. doi: 10.1007/s10459-012-9435-3

Mahasneh, O., Ayasrah, M., Yahyaa, S., and Al-maraziq, I. A. (2022). Educational implications of distance learning within the coronavirus pandemic (COVID-19) from the point of view of university students. *World J. Educ. Technol. Curr. Issues* 14, 655–670. doi: 10.18844/wjet.v14i3.7199

Mathivanan, S. K., Jayagopal, P., Ahmed, S., Manivannan, S. S., Kumar, P. J., Raja, K. T., et al. (2021). Adoption of e-learning during lockdown in India. *Int. J. Syst. Assur. Eng. Manag.*, 14, 1–10. doi: 10.1007/s13198-021-01072-4

Morreale, S. P., Thorpe, J., and Westwick, J. N. (2021). Online teaching: challenge or opportunity for communication education scholars? *Commun. Educ.* 70, 117–119. doi: 10.1080/03634523.2020.1811360

Nwabufo, B. N., Umoru, T. A., and Olukotun, J. O. (2013). The challenges of e-learning in tertiary institutions in Nigeria. In International conference the future of education Florence, June.

Pham, V. P. H., and Da Vo, T. N. (2021). CALL in Asia during Covid-19 and models of e-learning. In 17th international conference of the Asia Association of Computer-Assisted Language Learning (Asia CALL 2021) Atlantis Press.

Radha, R., Mahalakshmi, K., Kumar, V. S., and Saravanakumar, A. R. (2020). E-learning during lockdown of Covid-19 pandemic: a global perspective. *Int. J. Control Autom.* 13, 1088–1099. Available at: http://sersc.org/journals/index.php/IJCA/article/view/26035

Randazzo, M., Priefer, R., and Khamis-Dakwar, R. (2021). Project-based learning and traditional online teaching of research methods during COVID-19: an investigation of research self-efficacy and student satisfaction. *Front. Educ.* 6:662850. doi: 10.3389/feduc.2021.662850

Reavley, N. J., Morgan, A. J., Fischer, J. A., Kitchener, B., Bovopoulos, N., and Jorm, A. F. (2018). Effectiveness of eLearning and blended modes of delivery of mental health first aid training in the workplace: randomised controlled trial. *BMC Psychiatry* 18, 1–14. doi: 10.1186/s12888-018-1888-3

Reddy, P., Chaudhary, K., Sharma, B., and Chand, R. (2021). The two perfect scorers for technology acceptance. *Educ. Inf. Technol.* 26, 1505–1526. doi: 10.1007/s10639-020-10320-2

Reddy, P., Sharma, B. N., and Chaudhary, K. C. (2020). Measuring the digital competency of freshmen at a higher education institute. In Pacific Asia Conference on Information Systems. Association for Information Systems.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.* 66, 688–701. doi: 10.1037/h0037350

Saeed Al-Maroof, R., Alhumaid, K., and Salloum, S. (2020). The continuous intention to use e-learning, from two different perspectives. *Educ. Sci.* 11:6. doi: 10.3390/educsci11010006

Sharma, B. N., Nand, R., Naseem, M., Reddy, E., Narayan, S. S., and Reddy, K. (2018). Smart learning in the Pacific: design of new pedagogical tools. In 2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE) (pp. 573–580). IEEE.

Shofwan, I., Aminatun, S., Handoyo, E., and Kariadi, M. T. (2021). The effect of e-learning on Students' learning interest in the equivalence education program. *J. Nonform. Educ.* 7, 103–111. doi: 10.15294/jne.v7i1.29276

Sithole, A., Mupinga, D. M., Kibirige, J. S., Manyanga, F., and Bucklein, B. K. (2019). Expectations, challenges and suggestions for faculty teaching online courses in higher education. *Int. J. Online Pedag. Course Design* 9, 62–77. doi: 10.4018/IJOPCD.2019010105

Smith, G. G., Passmore, D., and Faught, T. (2009). The challenges of online nursing education. *Internet High. Educ.* 12, 98–103. doi: 10.1016/j.iheduc.2009.06.007

Tang, K. Y., Chang, C. Y., and Hwang, G. J. (2021). Trends in artificial intelligence-supported e-learning: a systematic review and co-citation network analysis (1998–2019). *Interact. Learn. Environ.*, 1–19. doi: 10.1080/10494820.2021.1875001 [ahead-of-print].

Tawarah, H., Mahasneh, O. M., and Al-Shuaybat, W. (2022). Factors affecting in achievement of universal courses objective by using distance education during COVID-19 pandemic. *J. Curric. Teach.* 11:43. doi: 10.5430/jct.v11n8p43