Adv Physiol Educ 47: 788–795, 2023.
First published August 24, 2023; doi:10.1152/advan.00257.2022

american physiological society

ADVANCES IN PHYSIOLOGY EDUCATION

EDUCATION RESEARCH

Remote vs. face-to-face activities in the teaching of renal pathophysiology in the context of social isolation during the early phase of the COVID-19 pandemic

Ahmed Haydar, Itamar Souza Santos, Luis Carlos Arcon, Mílton de Arruda Martins, Patricia Zen Tempski, and Roberto Zatz

Nephrology Division, Department of Clinical Medicine and Center for Development of Medical Education, School of Medicine, University of São Paulo, São Paulo, Brazil

Abstract

The advent of the COVID-19 pandemic forced medical schools around the world to adopt emergency remote learning as a resort to avoid interruption of courses. However, the effectiveness of online classes as an educational strategy has been questioned by medical educators and students. In a prospective observational study design, students enrolled in a renal physiology and pathophysiology course were exposed to either face-to-face or remote synchronous classes. Students taught online obtained significantly higher mean scores than the group who had in-person classes, both groups assessed with identical exams. Appropriate screening tests suggested that fraud is unlikely to have significantly influenced these results and that the observed differences in performance reflected increased learning by the remote group. These observations suggest that online classes can help to maintain the continuity of physiology and pathophysiology courses during periods of social isolation and may contribute to improving learning under normal conditions.

NEW & NOTEWORTHY In this study, we were able to make a rare direct comparison of face-to-face and remote strategies for the teaching of undergraduate medical students in a specific area, namely, renal pathophysiology. Unexpectedly, students who attended the remote course had significantly higher grades than those who had mostly in-person classes.

audiovisual aids; computer-assisted instruction; education; educational measurement; graduate

INTRODUCTION

The COVID-19 pandemic imposed restrictions on mobility, outdoor and indoor gatherings, as well as in-class activities (1). Medical education rapidly adapted to the new epidemiological reality: synchronous and asynchronous remote classes, as well as online assessments, took the place of in-person activities, leaving little room for practical face-to-face and "hands-on" educational strategies (2–4). As a natural consequence, mainstream technology-mediated teaching evolved to technology-exclusive teaching, complexifying the definitions of e-learner, e-teaching, and e-learning (5, 6). In this context, the adoption of new techniques to maintain students' routine, motivation, interaction, and collaboration has become, more than ever, a necessity (6–8). Indeed, several educational applications were developed, including virtual learning environments, online meeting systems, video communication and conferencing platforms, interactive polls and gamification presentations, as well as proctored and unproctored e-assessment software (3, 6–10).

In this unprecedented scenario, teachers around the globe were motivated to create, adapt, and incorporate new educational technologies quickly (3). Despite the innovations that sprouted, great uncertainty remains as to the efficacy of remote strategies in medical education (11–13). Students tend to perceive online teaching as less interactive and less effective for professional qualification (12). By contrast, a recent meta-analysis suggested that remote teaching may even be superior to face-to-face activities (14). In the same study, an important limitation was underlined: the scarcity of studies featuring direct comparisons between remote and face-to-face learning; such studies are hard to develop because of obvious ethical constraints.

In our institution, the challenges imposed by the emerging COVID-19 pandemic also led us to develop a prospective observational study, comparing in-person and remote strategies for the teaching of a specific area, namely, renal pathophysiology (15).

METHODS

In the School of Medicine of the University of São Paulo, the traditional Physiological Basis of Nephrology (PBN) core course is offered to third-year medical students. During 2 mo of the first semester, the following topics are discussed in 11 modules: Glomerular Filtration and Its Disorders; Renal Sodium Handling and Diuretics; Disorders of Water Balance; Mechanisms of Dehydration; Mechanisms of Edema; Mechanisms of Hypertension;

Correspondence: R. Zatz (r.zatz@fm.usp.br).
Submitted 11 January 2023 / Revised 24 July 2023 / Accepted 22 August 2023

Potassium Disorders; Calcium and Phosphorus Disorders; Fundamentals of Acid-Base Equilibrium; Acid-Base Disorders; and Mechanisms of Nephrolithiasis.

For each module, the educational strategy consists of a 1-h face-to-face lecture using PowerPoint animations as a tool to illustrate complex processes such as ion transport, urine acidification, and bone remodeling, followed by a session of hands-on small-group activities. In these "practical" classes, carried out in a room with 45 computer stations, the students are divided in groups of up to five and instructed to use software, developed locally with Excel, Visual Basic, and Delphi platforms, to change physiological variables such as fluid intake, acid production, and arteriolar resistances, observing instantaneously the resulting change of parameters such as urine output, blood pH, and glomerular pressure. These dynamic exercises always referred to the topics that were addressed during the respective lectures. The students were encouraged to perform virtual "experiments" and to discuss the results among themselves. In addition, students were allowed additional study time (8 h per week), predefined in their syllabus.

At the end of the course, students are assessed through a face-to-face proctored summative exam (SE), consisting of 50 multiple-choice questions (MCQs), with five alternatives and one right answer each. The students are given 3 h to complete the exam. Both the questions and their respective alternatives are arranged in random order. All questions are assigned the same weight. The final grade is calculated as simply the percentage of right answers, with no penalty imposed for wrong or blank answers. We adopt an open-book examination (OBE) format, based on the assumption that the future professionals will have access to a wealth of information and tools when dealing with real-life health care situations. Therefore, the exams focus on problem solving and critical knowledge, rather than on simple memorization. Accordingly, the MCQs are designed to assess higher levels of competency, such as applying knowledge in clinical scenarios, interpreting variations in physiological variables, analyzing physiological tests, and even proposing treatments. None of the required answers is readily obtained by consulting books or other materials. Students are not allowed to contact each other or seek advice from the teachers (16–19).

All these educational activities and assessment methods had to undergo urgent adaptation as of March 2020, as a consequence of the social restrictions adopted to contain the pandemic.

Study Design

About 170–180 students are admitted to this medical course once a year. Because of the large number of students, classes are divided into two groups, named *A* and *B*. Since this assignment is entirely random, these two groups are expectedly comparable.

From March to June 2020, *group A* (designated "control") had 8 out of the 11 classes of the PBN course in the traditional model and 3 remote, synchronous classes, with the 1-h PowerPoint lectures delivered with Google Meet. *Group B* (designated "intervention") began classes in April and attended all 11 classes remotely. Given the impossibility of keeping the "practical," hands-on activities in the remote context, the teacher instead demonstrated to the class the effects of varying selected renal parameters, exposing graphically the pathophysiological reasoning. Before each lecture, both groups were granted access to the respective slide presentation. In addition, both groups had unrestricted subsequent access to the video recordings of the remote classes (lectures and graphical demonstrations). Since such recordings were not performed before the pandemics, *group A* was able to watch the video recordings of the last three classes only.

The final assessment was based on the same type of 3-h 50-MCQ OBE applied under normal circumstances. In view of the circumstances, the exam was unproctored and held remotely, with Google Forms. Both groups sat the exam 11 days after the last class.

The differences between *groups A* and *B* regarding the teaching and assessment strategies adopted during the first semester of 2020 are summarized in Table 1.

All 173 students enrolled in the course in 2020 were invited to join the study. To participate, written informed consent was required. No incentive or direct advantage was offered. The study was approved by the Ethical Committee of the Faculty of Medicine of the University of São Paulo (467 1291 3 0000 0068).

The grades obtained in the SE, i.e., the proportion of correct answers in the 50-MCQ OBE, were utilized as a proxy for knowledge acquisition. The 2020 SE, focused on critical knowledge and problem solving, is given as Supplemental Material (available at https://doi.org/10.6084/m9.figshare.21861939.v2). Scores varied between 0 and 10. Hit rates (HRs) for each MCQ were defined as $N/T \times 100$, where N is the number of students answering the question correctly and T is the total number of students taking the test. A discrimination index (DI) was calculated as well. This index is intended to evaluate the question's ability to discern students with the best scores (HR of top 27%) from those with the worst ones (HR of bottom 27%). The DI is calculated as (HR top 27% − HR bottom 27%)/(HR top 27%) $\times$100. The questions of the 2020 SE with the top 10 DIs are given as Supplemental Material.

We compared the mean scores of *groups A* and *B* with those obtained in previous and in subsequent years, using pairwise Wilcoxon rank-sum tests and Holm *P*-value adjustment for multiple comparisons. Until 2020, all classes were held face to face (Table 2). Whereas in 2021 the PBN course was entirely remote, all classes were delivered in person in 2022. However, the SE, conceived and designed in the same way as described above, was remote and unproctored in both these years. In this manner, the performance of third-year students exposed to face-to-face (in 2022) and remote (in 2021) classes could be directly compared.

To assess whether possible differences in test performance could be ascribed to unsuspected inherent differences between groups, we developed a linear regression model to predict scores based on previous performance. The independent variables were the weighted mean of previous grades (WMPG), which is analogous to grade point average, and scores from two prerequisite courses [Kidney Anatomy/Physiology (KAP score) and Membrane Physiology (MP score)], all retrieved from medical school records (20–22). The corresponding equation is given as Supplemental Material. Based on this equation, it was possible to calculate the residuals, defined as the differences between the observed and predicted 2020 SE scores.

Table 1. *Summary description of groups and of teaching/assessment strategies*

Traditional	Control/*Group A*	Intervention/*Group B*
11 face-to-face classes	8 face-to-face classes	0 face-to-face classes
0 remote synchronous classes	3 remote synchronous classes	11 remote synchronous classes
Hands-on computer simulation classes in small groups, using computer stations	8 hands-on computer simulation classes in small groups, using computer stations	11 "practical" classes with demonstration of the effect of changing physiological variables
	3 "practical" classes with demonstration of the effect of changing physiological variables	
Classes not recorded	3 recorded remote classes available asynchronously.	11 recorded remote classes available asynchronously.
Face-to-face proctored open-book examination	Remote unproctored open-book examination	
50 multiple-choice questions	50 multiple-choice questions	
5 alternatives, 1 best answer	5 alternatives, 1 best answer	

We also calculated the coefficient of determination (R^2) value for the model, as a measurement of the variation of 2020 SE scores that can be explained by variation of the WMPG and the KAP score (23). If the latter explains the differences found in the 2020 PBN test scores, the residuals are expected to have a similar distribution in *groups 2020/A* and *2020/B*. Deviations from this behavior could only be ascribed to differences in variables that are external to the model, such as the adapted teaching methods utilized in 2020.

Differences in the performances of *groups A* and *B* could simply reflect fraud, rather than any disadvantages (or advantages) brought about by the forced adoption of remote classes. Considering the remote and unproctored context in which the e-assessment was carried, the possibility of cheating during the exam cannot be ignored. It is also conceivable that students of the intervention group, who took the test 2 mo later, might have been tipped as to its content by colleagues in the control group. To assess whether fraud could account for possible differences between groups, we used the U3 person fit statistics as proposed by Van Der Flier (20, 24). With the U3 technique it is possible to analyze responses and identify the degree to which they deviate from an expected pattern, based on the deterministic Guttman model. Briefly, this model assumes that the ideal pattern of answers for a student who has obtained, for instance, an 8.0 score (on a 0–10 scale) should be that his/her right answers concentrate among the 80% easiest questions, while the 20% hardest ones are missed. The U3 statistic measures the extent to which the response patterns diverge from the Guttman model: deviance is rated from 0 to 1, with 0 representing exact fitting to and 1 representing maximal departure from the model. In the example of the student with the 8.0 score, a U3 of 1 means that 80% of the hardest questions were hit, while the 20% easiest ones were missed. This could be explained by lucky guessing or, more likely, by cheating (20–22, 25, 26).

Categorical variables were expressed as proportions, with corresponding 95% confidence intervals (95% CIs). Chi-square and Fisher's exact tests were used whenever appropriate. Since the continuous variables were not normally distributed, differences between *groups A* and *B* were assessed by Wilcoxon rank-sum tests. All analyses were performed with R software and the PerFit R package (27, 28). Significance level was set at 0.05.

RESULTS

One-hundred fifty-one students agreed to participate. Of the 22 students who did not provide written informed consent, none declared the reasons for the refusal. Five students were excluded because of missing test data (failed to answer 1 or more items), leaving 146 students for the final analysis (81 from *group A* and 65 from *group B*). To the best of our knowledge, no selection bias was introduced as a result of missing data or refusal to participate, as grades were similar between included and not included students (data not shown).

Scores obtained by *groups A* and *B* in the PBN SE are shown in Fig. 1*A*. The median score obtained by *group A* was 7.8 (interquartile range 7.0–8.4). *Group B*'s score was significantly higher (median 8.6, interquartile range 8.2–9.0; $P < 0.001$). *Group B* obtained a significantly higher mean HR in 19 of 50 MCQ test items. Of these, the HR difference was higher than 30% in questions with the highest DI.

A slight but significant difference in the WMPG score was observed between *groups A* and *B* (7.88 and 8.06,

Table 2. *Adjusted P values for each paired comparison*

	2018A	2018B	2019A	2019B	2020A	2020B	2021A	2021B	2022A
2018B	1.000								
2019A	0.201	0.392							
2019B	1.000	1.000	0.467						
2020A	0.467	1.000	1.000	1.000					
2020B	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**				
2021A	**<0.001**	**<0.001**	1.000	**<0.001**	0.171	**<0.001**			
2021B	**0.001**	**0.001**	1.000	**<0.001**	0.149	**0.045**	1.000		
2022A	1.000	1.000	**0.018**	1.000	0.063	**<0.001**	**<0.001**	**<0.001**	
2022B	1.000	1.000	0.489	1.000	1.000	**<0.001**	**<0.001**	**<0.001**	1.000

Significant differences are in bold.

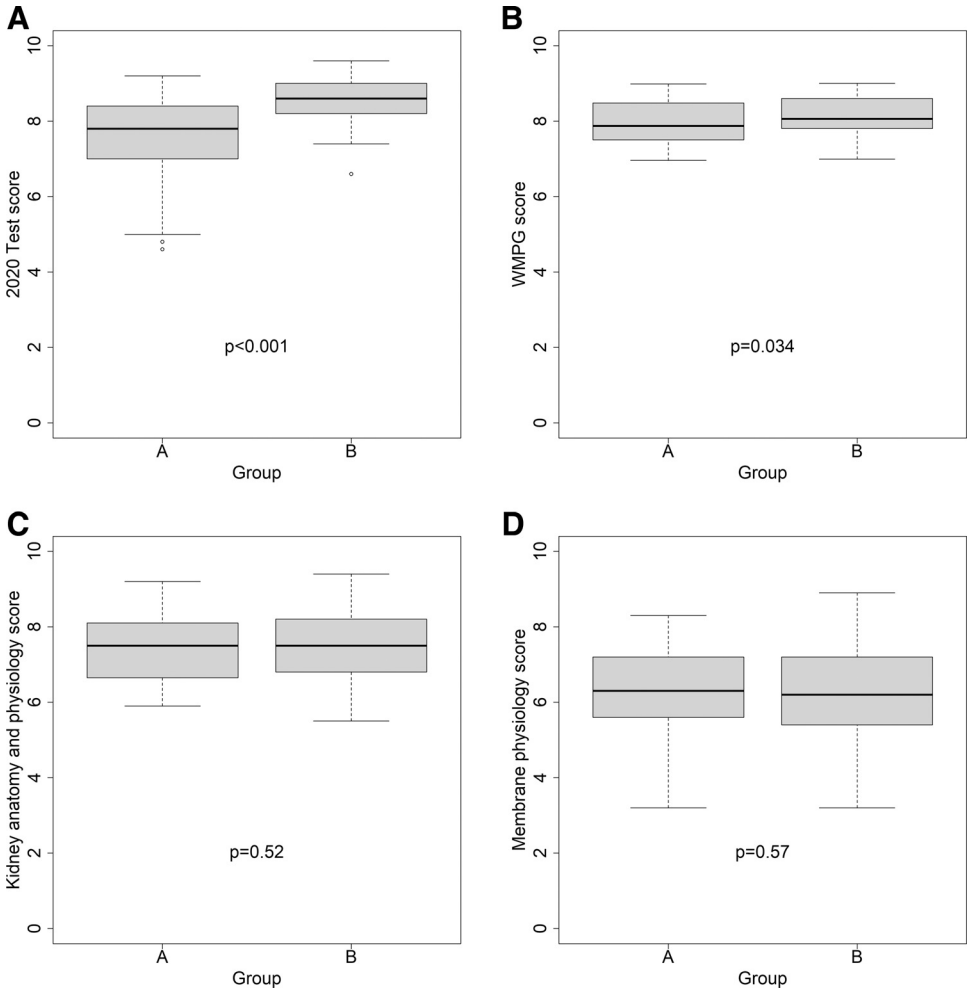

Figure 1. Score distributions for the study sample. *A*: in the 2020 Physiological Basis of Nephrology (PBN) test. *B*: weighted mean of previous grades (WMPG). *C*: Kidney Anatomy/Physiology. *D*: Membrane Physiology.

respectively; $P = 0.034$) (Fig. 1*B*). Comparison of grades obtained by *groups A* and *B* in prerequisite courses (Kidney Anatomy and Physiology and Membrane Physiology) showed no significant differences (Fig. 1, *C* and *D*, respectively).

Only two students could not be included in the linear regression model, because their predicting grades were unavailable. The R^2 coefficient, which measures how accurately the independent variables (WMPG, KAP and MP scores) predict the PBN test grades, was 0.155. This indicates that previous performances explained only 15.5% of the variance of the PBN exam scores. This first evidence supports our hypothesis that the groups are not inherently different.

The distribution of residuals is shown in Fig. 2. The mean residual for *group B* was significantly higher ($P < 0.001$) than that observed for *group A*, corroborating the notion that the better performance of *group B* in the SE is not due to preexisting differences between groups.

Figure 3 shows the hits (light rectangles) and misses (dark rectangles) registered for each student (organized in rows) and each question (organized in columns). Students are ordered from the lowest (*bottom*) to the highest (*top*) grades. Test items are organized from the easiest (*left*) to the hardest (*right*) questions. The continuous black line in both panels of Fig. 3 indicates, according to each student's grade, which questions he/she was expected to hit (left side of the line) or miss (right side of the line). In this manner, it is possible to visualize whether the performance of each student deviates from the expected pattern of hitting the easiest questions and missing the hardest ones. This representation shows clearly that *groups A* and *B* behaved similarly and that, if anything, *group B*'s performance was closer than *group A*'s to that expected of an ideal noncheating class. *Group B*'s mean U3 statistic was significantly lower than that observed in *group A* (0.121 and 0.162, respectively, $P = 0.003$). These observations argue strongly against the hypothesis that the better performance of *group B* in the summation test was due to fraud.

Attendance was higher in remote classes. In the first eight (on site) classes, *group A*'s mean attendance was 87.0 ± 7.7%. During the last three (remote) classes, *group A*'s mean attendance (93.9 ± 1.6%) was similar to that of *group B* (93.2 ± 4.8%), which had online classes only.

Mean grades obtained by *groups A* and *B* during the first semester of 2020 (excluding PBN) were 8.1 ± 0.6 and 8.3 ± 0.6, respectively ($P > 0.05$). Accordingly, third-year students of *groups A* and *B* obtained similar mean grades in the PBN course in 2021 (all classes synchronous and remote). Of note, these grades were significantly higher than in 2018 and 2019 (except for *group B*'s grades in 2020). Additionally, after return to face-to-face classes in 2022, with the SE still in the

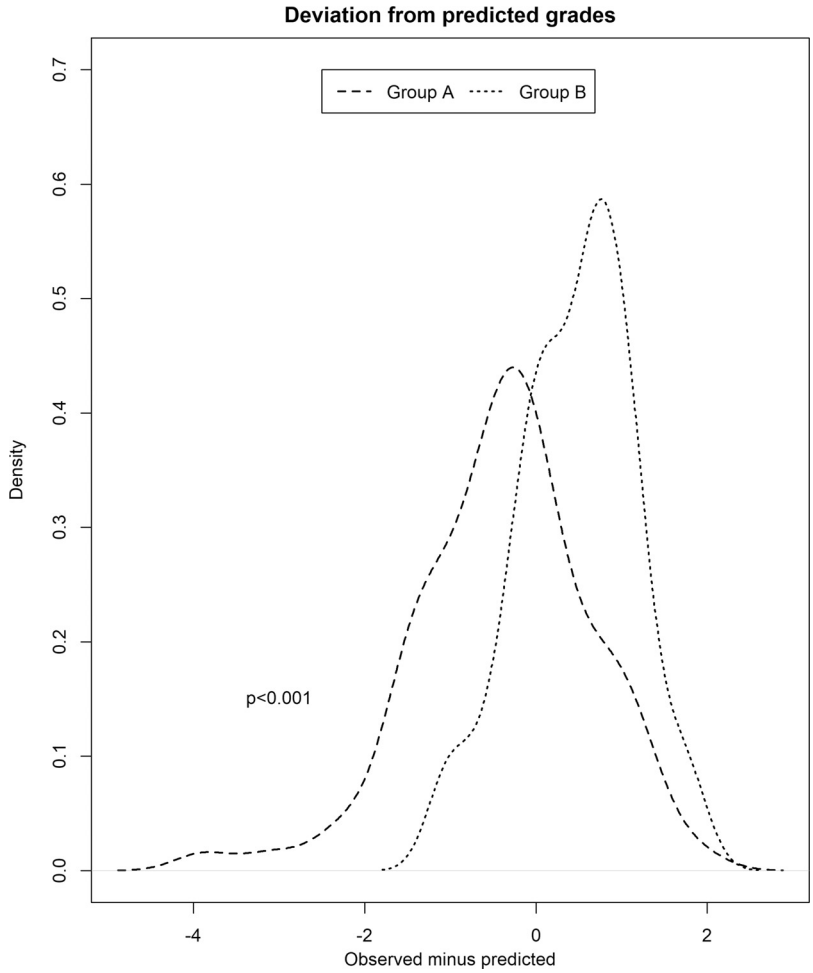

Figure 2. Distribution of residuals in the summative 2020 Physiological Basis of Nephrology (PBN) test (actual minus predicted scores) for *groups A* and *B*.

unproctored OBE format, grades returned to levels similar to those obtained by *group A* in 2020 and by both groups in previous years (Fig. 4).

In a post hoc comparison of the performance of the same students in Clinical Nephrology (in 2021, when these students were in the fourth year), *groups A* and *B* obtained mean grades of 8.1 ± 1.4 and 8.1 ± 1.6, respectively ($P > 0.05$).

DISCUSSION

Social restrictions imposed by the COVID-19 pandemic created an unusual scenario: half of the 2020 third-year medical students had the majority of the PBN course in a conventional face-to-face fashion (*group A*), whereas *group B* was exposed to an exclusively remote educational strategy with synchronous

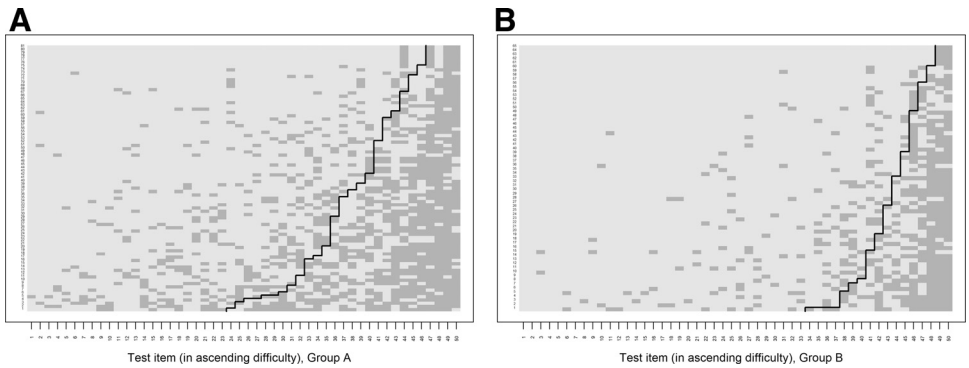

Figure 3. Hit-and-miss plot for the study sample in the 2020 Physiological Basis of Nephrology (PBN) test. *A*: *group A*. *B*: *group B*. In these diagrams, each row represents an individual student, whereas each column represents 1 of the 50 multiple-choice questions (MCQs) constituting the summative exam. Hits are represented by the light rectangles and misses by the dark rectangles. For each student (rows) the short vertical black line indicates the total number of hits (and therefore the score) obtained by that particular student. The distribution of hits and misses indicates that both groups behaved as would be expected in the absence of cheating, with hits tending to concentrate to the left, and misses to the right, of these vertical lines, which are joined to constitute a single line, thus facilitating visualization.

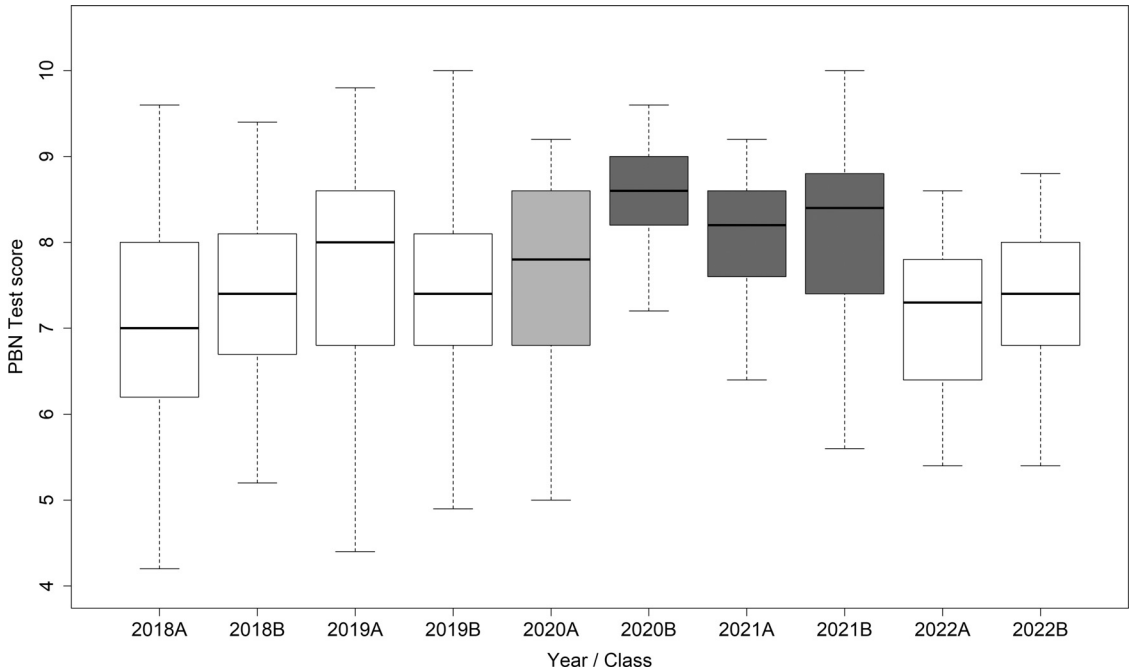

Figure 4. Physiological Basis of Nephrology (PBN) test scores from 2018 to 2022. In this graph, each box plot represents mean test scores for *groups A* and *B* in the respective year. The distance learning groups (*2020B*, *2021A*, and *2021B*, represented in gray) had considerably higher mean scores compared with the classes of 2018, 2019, 2020A, and 2022 ($P < 0.001$). *Group 2020A*, which had only 3 in-presence classes, is represented with a lighter shade of gray.

classes. This context allowed us to design a prospective observational study, with direct comparison between the two schemes.

We hypothesized initially that remote teaching would prove inferior to conventional educational strategies, which are expected to allow direct interaction of students with both teachers and colleagues in lectures and practical activities. Unexpectedly, *group B*'s scores in the summative exams were significantly higher than *group A*'s and PBN historical mean grades. The PBN course had followed the same model since 2018. From 2018 until *group A*'s course in 2020, classes had a relatively constant performance in the SE. The redesigned course offered to the *group B* in 2020 was maintained in 2021, now for *groups A* and *B*, both of whom exhibited better performances than in previous years, albeit with slightly lower grades than those of *group B* in 2020. Interestingly, when in 2022 PBN returned to the conventional face-to-face model, grades obtained by both *group A* and *group B* third-year students fell to the levels obtained in 2018, 2019, and (in *group A*) 2020.

The better performance of *group B* in the first semester of 2020 could indicate that, despite the random distribution, this group was intrinsically more competent than *group A*. This explanation becomes less likely when we consider that *1*) the previous difference between groups regarding WMPG, though significant, was trivial (slightly above 2%); *2*) variations in WMPG and prerequisites explain only 15.5% of the test scoring dispersion; *3*) the distribution of residuals (observed minus predicted scores) in *group B* was distinctively shifted to the right, indicating that because of the intervention of some new factor the scores obtained by this group were higher than would be expected by its retrospect; and *4*) no grade differences were seen in other courses offered during the first semester of 2020.

Another explanation for the observed differences could be the occurrence of fraud. Given the unproctored nature of the SE, it is conceivable that students might have communicated with each other, sharing the resolution of the tests, especially the most difficult ones. Additionally, since *group B* took the exam 2 mo later, students of this group might have been favored by previous knowledge of the tests provided by colleagues in *group A*. However, these possibilities are not supported by the U3 statistics, since *group B* had lower deviance scores and a pattern of hits and misses that, if anything, was closer to that expected for noncheating students compared with *group A*. Moreover, it must be stressed that in both groups U3 statistics were relatively low, which is inconsistent with the existence of a significant degree of cheating.

Although this aspect could not be evaluated in the present study, the better performance of *group B* in 2020 might be linked to the impact of social isolation, with time previously spent in commuting and extracurricular activities now channeled into studying and attending remote classes and even social activities. In addition, at least part of the students may have felt protected, under social isolation, from the risk of becoming ill (4, 29). On the other hand, social isolation may have exerted a negative impact as well. In a specific study of our institution, students reported feeling anxious and insecure with the sudden change, fearing not having an adequate formation. In addition, they deplored the loss of social interaction and community daily life, considered essential to the development of professional identity (30). Thus, the real impact of the social isolation imposed by the pandemic is uncertain at this time.

Since the superior performance of *group B* cannot be ascribed to either preexistent differences between groups or fraud, it seems reasonable to conclude that the differences

between groups reflect the distinct educational strategies to which the groups were exposed. The reasons for this apparently higher efficiency of online classes are unclear and warrant further investigation. Students attended classes more regularly and may have found fewer reasons for distraction when individually focused on online activities, especially during the (forcefully) expositive "practical classes," when the concepts addressed in the lectures are reviewed under a different angle and solidified, without collateral talk. Another possible explanation is the different exposure of *groups A* (3) and *B* (11) to the recorded classes, although we were unable to quantify the utilization of this resource or to assess the individualization of the learning tracks (4, 13). In addition, it should be noted that students of both groups had access to all the slide presentations. Thus, the role of the recorded lectures is uncertain. A final possibility is that the computer simulations such as those utilized in the PBN course work better as a teaching tool in a remote learning context, although evidence in favor of this concept is presently lacking. At any rate, it is not possible at this time to ascertain whether the apparent superiority of remote classes would still be observed if they were part of a planned teaching strategy rather than adopted as an emergency measure in the context of social isolation.

Whatever the mechanisms underlying the higher grades obtained by *group B*, the knowledge acquired in the 2020 PBN was not reflected in a better subsequent performance in 2021 Clinical Nephrology (CN), since no difference was seen in the mean grades obtained by these students in that discipline, which is offered in the fourth year of medical school. This finding should be interpreted with caution, however, since the CN course is heavily based on anamnesis and examination of hospitalized patients under investigation for renal disease. Moreover, the students are assessed differently in the CN course, with a closed-book, open-ended, written SE. At any rate, the similarity of the grades obtained in CN is consistent with the notion that, in the long run, remote teaching is noninferior to face-to-face classes.

Remote learning cannot permanently and completely replace face-to-face activities, which are essential to build teacher-student and student-student relationships and develop discipline and self-regulation skills, communication, and teamwork competencies (4, 13, 31). Nevertheless, online activities may represent an important complement to traditional classes, with blended systems, or even exclusively online classes, finding their way into the future teaching of renal physiology and pathophysiology (30).

Limitations

This study adopted an observational design, with limitations due to the impossibility of directly randomizing students to either method. In addition, the study assumes the idea of a closed system: any change in learning outcome derives directly from applied interventions (32). This reductionist conception, based on linearity and causality, may be questioned in view of the variety and complexity of the available educational systems and of the factors involved in the process of learning itself (32, 33).

Another possible limitation is the highly contextualized nature of the present findings: they may not be applicable outside the setting of a basic science discipline taught in a Brazilian medical school amidst a raging pandemic. In a properly preorganized remote learning environment, in opposition to the emergency remote scenario, results may be different (34).

Moreover, the present study was focused on the assessment of knowledge, with no comparison regarding the acquisition or development of attitudes and skills. The use of exam scores as a proxy for knowledge is a simplistic way of evaluating the whole learning process. Although limited, the data correspond to the second level of Kirkpatrick's model, which is a widespread evaluation framework. In addition, scores correspond to an easily reproducible outcome that allows comparison with other studies (15, 35).

Finally, we could not establish the long-term impact of the remote learning strategy. The similar grades obtained by *groups A* and *B* in a subsequent course suggest noninferiority of remote classes.

Conclusions

In this study, the adoption of an online synchronous educational strategy appeared to be superior to conventional face-to-face activities. Multiple analyses indicated that fraud is unlikely as a relevant factor, suggesting that the difference in performance between groups may indeed reflect more efficient learning.

By directly comparing face-to-face and remote learning, this study may contribute significantly to the development of strategies for medical education. Although the present results are not to be construed as unequivocal evidence that online classes are superior to in-person activities, this method may become a valuable ancillary tool in the future teaching of renal physiology, and possibly other disciplines as well.

DATA AVAILABILITY

Data will be made available upon reasonable request.

SUPPLEMENTAL MATERIAL

Supplemental Material: https://doi.org/10.6084/m9.figshare.21861939.v2.

ACKNOWLEDGMENTS

We are grateful to Giovana Cardoso Machado Boer for superb technical support with the online classes.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the authors.

AUTHOR CONTRIBUTIONS

R.Z. conceived and designed research; L.C.A. and R.Z. performed experiments; A.H., I.S.S., and R.Z. analyzed data; A.H., I.S.S., and R.Z. interpreted results of experiments; I.S.S. and R.Z. prepared figures; A.H. and R.Z. drafted manuscript; A.H., M.d.M., P.Z.T., and R.Z. edited and revised manuscript; A.H., M.d.M., P.Z.T., and R.Z. approved final version of manuscript.

REFERENCES

1. **del Rio C**, **Malani PN.** 2019 novel coronavirus—important information for clinicians. *JAMA* 323: 1039–1040, 2020. doi:10.1001/jama.2020.1490.
2. **Rose S.** Medical student education in the time of COVID-19. *JAMA* 323: 2131–2132, 2020. doi:10.1001/jama.2020.5227.
3. **Gordon CM**, **Ackerman KE**, **Berga SL**, **Kaplan JR**, **Mastorakos G**, **Misra M**, **Murad MH**, **Santoro NF**, **Warren MP.** Functional hypothalamic amenorrhea: an Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab* 102: 1413–1439, 2017. doi:10.1210/jc.2017-00131.
4. **Tempski P**, **Danila AH**, **Arantes-Costa FM**, **Siqueira MA**, **Torsani MB**, **Martins MA.** The COVID-19 pandemic: time for medical teachers and students to overcome grief. *Clinics (Sao Paulo)* 75: e2206, 2020. doi:10.6061/clinics/2020/e2206.
5. **Daniel J.** Making sense of MOOCs: musings in a maze of myth, paradox and possibility. *J Interact Media Educ* 2012: 18, 2012. doi:10.5334/2012-18.
6. **Ellaway R**, **Masters K.** AMEE Guide 32: e-learning in medical education. Part 1: Learning, teaching and assessment. *Med Teach* 30: 455–473, 2008. doi:10.1080/01421590802108331.
7. **Martin F**, **Bolliger DU.** Engagement matters: student perceptions on the importance of engagement strategies in the online learning environment. *OLJ* 22: 1, 2018. doi:10.24059/olj.v22i1.1092.
8. **Kaup S**, **Jain R**, **Shivalli S**, **Pandey S**, **Kaup S.** Sustaining academics during COVID-19 pandemic: the role of online teaching-learning. *Indian J Ophthalmol* 68: 1220–1221, 2020. doi:10.4103/ijo.IJO_1241_20.
9. **Banna J**, **Lin MF**, **Stewart M**, **Fialkowski MK.** Interaction matters: strategies to promote engaged learning in an online introductory nutrition course. *J Online Learn Teach* 11: 249–261, 2015.
10. **Choules AP.** The use of elearning in medical education: a review of the current situation. *Postgrad Med J* 83: 212–216, 2007. doi:10.1136/pgmj.2006.054189.
11. **Monaghan AM.** Medical teaching and assessment in the era of COVID-19. *J Med Educ Curric Dev* 7: 2382120520965255, 2020. doi:10.1177/2382120520965255.
12. **Dost S**, **Hossain A**, **Shehab M**, **Abdelwahed A**, **Al-Nusair L.** Perceptions of medical students towards online teaching during the COVID-19 pandemic: a national cross-sectional survey of 2721 UK medical students. *BMJ Open* 10: e042378, 2020. doi:10.1136/bmjopen-2020-042378.
13. **Stoehr F**, **Müller L**, **Brady A**, **Trilla A**, **Mähringer-Kunz A**, **Hahn F**, **Düber C**, **Becker N**, **Wörns MA**, **Chapiro J**, **Hinrichs JB**, **Akata D**, **Ellmann S**, **Huisman M**, **Koff D**, **Brinkmann S**, **Bamberg F**, **Zimmermann O**, **Traikova NI**, **Marquardt JU**, **Chang D-H**, **Rengier F**, **Auer TA**, **Emrich T**, **Muehler F**, **Schmidberger H**, **Baeßler B**, **Dos Santos DP**, **Kloeckner R.** How COVID-19 kick-started online learning in medical education—the DigiMed study. *PLoS One* 16: e0257394, 2021. doi:10.1371/journal.pone.0257394.
14. **Pei L**, **Wu H.** Does online learning work better than offline learning in undergraduate medical education? A systematic review and meta-analysis. *Med Educ Online* 24: 1666538, 2019. doi:10.1080/10872981.2019.1666538.
15. **Frye AW**, **Hemmer PA.** Program evaluation models and related theories: AMEE Guide No. 67. *Med Teach* 34: e288–e299, 2012. doi:10.3109/0142159X.2012.668637.
16. **Feller M.** Open-book testing and education for the future. *Stud Educ Eval* 20: 235–238, 1994. doi:10.1016/0191-491X(94)90010-8.
17. **Brightwell R**, **Daniel JH**, **Stewart A.** Evaluation: is an open book examination easier? *Biosci Educ* 3: 1–10, 2004. doi:10.3108/beej.2004.03000004.
18. **Bloom BS**, **Airasian P**, **Cruikshank K**, **Mayer R**, **Pintrich P**, **Raths J.** *A Taxonomy for Learning, Teaching, and Assessing: a Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman, 2001.
19. **Durning SJ**, **Dong T**, **Ratcliffe T**, **Schuwirth L**, **Artino AR**, **Boulet JR**, **Eva K.** Comparing open-book and closed-book examinations. *Acad Med* 91: 583–599, 2016. doi:10.1097/acm.0000000000000977.
20. **Ranger J**, **Schmidt N**, **Wolgast A.** The detection of cheating on e-exams in higher education—the performance of several old and some new indicators. *Front Psychol* 11: 568825, 2020. doi:10.3389/fpsyg.2020.568825.
21. **Harmon OR**, **Lambrinos J.** Are online exams an invitation to cheat? *J Econ Educ* 39: 116–125, 2008. doi:10.3200/JECE.39.2.116-125.
22. **Dendir S**, **Stockton Maxwell R.** Cheating in online courses: evidence from online proctoring. *Comput Hum Behav Rep* 2: 100033, 2020. doi:10.1016/j.chbr.2020.100033.
23. **Lunt M.** Introduction to statistical modelling: linear regression. *Rheumatology (Oxford)* 54: 1137–1140, 2015. doi:10.1093/rheumatology/ket146.
24. **Van Der Flier H.** Deviant response patterns and comparability of test scores. *J Cross Cult Psychol* 13: 267–298, 1982. doi:10.1177/0022002182013003001.
25. **Dimitrov DM**, **Atanasov DV**, **Luo Y.** Person-fit assessment under the D-scoring method. *Measurement* 18: 111–123, 2020. doi:10.1080/15366367.2020.1725733.
26. **Haladyna TM**, **Rodriguez MC.** *Developing and Validating Test Items*. New York: Routledge, 2013.
27. **Tendeiro JN**, **Meijer RR**, **Niessen AS.** PerFit: an R package for person-fit analysis in IRT. *J Stat Soft* 74: 1–27, 2016. doi:10.18637/jss.v074.i05.
28. **R Development Core Team**. R: a language and environment for statistical computing. 2013. http://r.meteo.uni.wroc.pl/web/packages/dpIR/vignettes/intro-dpIR.pdf.
29. **Tempski P**, **Arantes-Costa FM**, **Kobayasi R**, **Siqueira MA**, **Torsani MB**, **Amaro BQ**, **Nascimento ME**, **Siqueira SL**, **Santos IS**, **Martins MA.** Medical students' perceptions and motivations during the COVID-19 pandemic. *PLoS One* 16: e0248627, 2021. doi:10.1371/journal.pone.0248627.
30. **Chinelatto LA**, **da Costa TR**, **Medeiros VM**, **Boog GH**, **Hojaij FC**, **Tempski PZ**, **Martins MA.** What you gain and what you lose in COVID-19: perception of medical students on their education. *Clinics (Sao Paulo)* 75: e2133, 2020. doi:10.6061/clinics/2020/e2133.
31. **Rowe M**, **Frantz J**, **Bozalek V.** The role of blended learning in the clinical education of healthcare students: a systematic review. *Med Teach* 34: e216–e221, 2012. doi:10.3109/0142159X.2012.642831.
32. **Geyer R**, **Mackintosh A**, **Lehmann K.** *Integrating UK and European Social Policy: the Complexity of Europeanisation*. Abingdon, UK: Radcliffe Publishing, 2005.
33. **Regmi K**, **Jones L.** A systematic review of the factors—enablers and barriers—affecting e-learning in health sciences education. *BMC Med Educ* 20: 91, 2020. doi:10.1186/s12909-020-02007-6.
34. **Hodges CB**, **Moore S**, **Lockee BB**, **Trust T**, **Bond MA.** The difference between emergency remote teaching and online learning. *EDUCAUSE Review*. 2020. https://vtechworks.lib.vt.edu/handle/10919/104648 [2023 Jun 25].
35. **Kirkpatrick JD**, **Kirkpatrick WK.** *Kirkpatrick's Four Levels of Training Evaluation*. Alexandria, VA: Association for Talent Development; 2016.