Research Article

Effects of COVID-19 Pandemic on Progress Test Performance in German-Speaking Countries

Victoria Sehy [ID],[1] Iván Roselló Atanet [ID],[1] Miriam Sieg [ID],[2] Jana Struzena,[1] and Maren März [ID][1]

[1]*Charité—Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, Progress Test Medizin, Charitéplatz 1, Berlin 10117, Germany*
[2]*Charité—Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, Progress Test Medizin and BIH QUEST Center for Transforming Biomedical Research, Translational Research Unit, Charitéplatz 1, Berlin 10117, Germany*

Correspondence should be addressed to Victoria Sehy; victoria.sehy@charite.de

Background. The COVID-19 pandemic has been the source of many challenges for medical students worldwide. The authors examined short-term effects on the knowledge gain of medical students in German-speaking countries. *Methods*. The development of the knowledge gain of medical students during the pandemic was measured by comparing the outcomes of shared questions within Berlin Progress Test (PT) pairs. The PT is a formative test of 200 multiple choice questions at the graduate level, which provides feedback to students on knowledge and knowledge gain during their course of study. It is provided to about 11,000 students in Germany and Austria around the beginning of each semester. We analyzed three successive test pairs: PT36-PT41 (both conducted before the pandemic), PT37-PT42 (PT37 took place before the pandemic; PT42 was conducted from April 2020 onwards), and PT38-PT43 (PT38 was administered before the pandemic; PT43 started in November 2020). The authors used mixed-effect regression models and compared the absolute variations in the percentage of correct answers per subject. *Results*. The most recent test of each PT pair showed a higher mean score compared to the previous test in the same pair (PT36-PT41 : 2.53 (95% CI: 1.31–3.75), PT37-PT42 : 3.72 (2.57–4.88), and PT38-PT43 : 5.66 (4.63–6.69)). Analogously, an increase in the share of correct answers was observed for most medical disciplines, with Epidemiology showing the most remarkable upsurge. *Conclusions*. Overall, PT performance improved during the pandemic, which we take as an indication that the sudden shift to online learning did not have a negative effect on the knowledge gain of students. We consider that these results may be helpful in advancing innovative approaches to medical education.

1. Introduction

The COVID-19 pandemic has impacted almost every area of daily life worldwide; students have also been affected by changes in their studies due to lockdown measures [1]. Undergraduate curriculums in medical schools usually include an extensive practical component implying regular contact with patients; students would therefore be put at risk for potential infection if practical lessons were to be held as they were planned before the pandemic [2]. The impact of all these circumstances on the academic performance of medical students has been the subject of research, with differing outcomes: there are studies reporting that student performance worsened [3, 4]; stagnated [5] or just changed for specific subjects [6] during the pandemic.

However, there might still be a knowledge gap to fill due to the methodological limitations in the literature published so far, among which we can mention small sample sizes, limited research scopes (i.e., only specific subjects or semesters were considered), or incomplete information about hypothetical differences in the difficulty levels of the exams being compared.

In order to help fill this knowledge gap, we set ourselves to investigate short-term effects on knowledge gain using data from recent issues of the Progress Test Medicine (PT).

The PT is a formative test including 200 multiple choice questions at the graduate level, which provides feedback to students on knowledge and knowledge gain during their course of study [7]. It is usually administered around the beginning of each semester, and in the summer of 2020, it was provided to 11,101 students from 15 German and Austrian faculties.

In addition to the large amount of data available and the possibility of observing the development in every semester and subject, another significant strength of the PT lies in the fact that it assesses current levels of knowledge without giving students the chance to prepare for it [7].

In summary, we intend to address the following two main questions throughout this study using data from PT:

(i) Is there a substantial change in knowledge between tests that took place prior to the pandemic ("pre-pandemic") and those conducted after the pandemic began?

(ii) Are there differences at the specialty level? A relevant question here is whether the observed changes were similar for all fields of study or if there were remarkable differences in performance depending on the medical discipline considered.

2. Materials and Methods

2.1. Setting. The PT is a low-stakes test on medical knowledge assembled every semester by Char-ité–Universitätsmedizin Berlin. PT exam regulations may differ between participating faculties: for example, participating in the test is mandatory for all students in 12 faculties and voluntary in the remaining two, and the number of compulsory participations demanded from students also varies by faculty. Additionally, three different test modalities are implemented depending on the faculty where the test is carried out: the traditional paper-pencil exam (which, however, has been completely abandoned because of the pandemic), the consortium's own platform (ePT), where students are required to state how confident they are that their answer is indeed true (ePT) [8] and also other learning platforms (e.g., ILIAS [9]) where this confidence statement is not included.

The "winter semester PT" takes place usually from October to December, while the "summer semester PT" is conducted from April to June. Since October 2019, our PTs share a considerable amount of questions with the PT that took place five semesters before. This leads to a natural pairing of tests administered five semesters (two and a half years) apart from each other. We conducted our analysis based on the shared questions within each of the pairs. Three consecutive PT pairs were included in our study; the first one comprises PT number 36 and PT number 41 (which in the following will be called "PT36" and "PT41"), which share 122 questions. As both tests took place before the pandemic, starting in April 2017 and October 2019, respectively, we included this dataset as a control. PT37 and PT42 share 155 questions; PT37 started before the pandemic in October 2017 and PT42 in April 2020, during the first lockdown in

Germany and Austria. Because the pandemic began to spread across Europe just a few weeks before the summer semester of 2020 was scheduled to start, new entrants in medical schools were suddenly confronted with a rather uncertain academic situation, having to adapt themselves to a completely virtual study environment, which had never been implemented on a comparable scale up to that time.

PT38 started in April 2018 and PT43 in November 2020, sharing 134 questions. By November 2020, online lectures had already become the norm, while practical lessons had been reduced or cancelled in line with mandatory social distancing regulations. Examination periods were postponed and prolonged, and sometimes more lenient rating procedures were applied, counting failed exams as "free shot."

Since both tests belonging to the same pair share most of their questions, students who took both tests in a pair were shown the shared questions twice, while the rest were presented these questions only once. To quantify the effect this might have on test results, we performed a t-test comparing both groups ("seen twice" vs. "seen once"). We estimated the effect size using Cohen's d; significance level α was set to $p = 0.01$. Students in the "seen twice" group did not outperform those in the "seen once" group (t-statistic -0.32, p-value 0.75 and effect size (Cohen's d) -0.01 for pair 36–41; 1.96, 0.05, 0.07 for pair 37–42, and 2.43, 0.02, 0.09 for pair 38–43, respectively).

2.2. Participants. A total of 9 faculties were included in this study; in addition to consenting to the use of their data, they had to meet two further requirements:

(i) Faculty-specific PT exam regulations must not have changed between the summer semester of 2017 and the winter semester of 2020

(ii) The faculty must have administered the test every academic semester since the summer semester of 2017

We used a pseudonymized dataset with the shared questions of PT36 and PT41, PT37 and PT42, and PT38 and PT43. These datasets contained the answers of participants to each question as well as the semester to which they belong, the pseudonymized faculty where they study, and whether their participation in the test is considered "serious" or not; participants classified as nonserious are excluded from the calculation of comparison groups since the validity of results would otherwise be jeopardized [10].

Nonserious participation is presumed when one or more of the following happens:

(i) The amount of time devoted to completing the test is too short [11] (less than 20 minutes)

(ii) Every single question of the test was either answered with "don´t know" or not answered at all [12]

(iii) None of the 120 last questions of the test were responded to (suggesting that the test was left incomplete with more than half of it yet to be read)

(iv) The self-monitoring accuracy rate in relation to testing answers is lower than 33% upon 20 or more questions, hinting that most answers were guessed or randomly chosen

In addition to disengaged test taking, exam misconduct (e.g., "cheating") must also be addressed as a construct-irrelevant factor. Two particular design elements of the PT are key to preventing exam misconduct: firstly, the test is purely formative and not linked to any specific course content. Besides, test scores have no effect whatsoever on the final grades of students [13,14]. Secondly, the timeframe for the test is tight (180 minutes for 200 questions) with no possibility of interruption [15].

2.3. Data Analysis

2.3.1. Overall Test Performance. For each PT pair, we fitted a linear mixed-effect model with random intercept and slope and with the relative PT score (correctly answered questions in percent) as the outcome variable. Four fixed effects predictors were set: test number, semester of study, the interaction between test number and semester of study, and test modality (digital vs. ePT vs. PT with paper-pencil). The medical school where each test was administered (random intercept) and the interaction between faculty and semester of study (random slope) were chosen as random effect predictors. This choice of predictors is based on the assumption (corroborated by long-term data from Progress Test) that curricular differences might lead to dissimilar semestral variations in average scores among the participating medical schools. Coefficients were fit with the restricted maximum likelihood approach.

We used R for Windows, version 3.6.1 (R Core Team, Vienna, Austria), and the lmer function from the R-package lme4 [16] for fitting the linear mixed models; additionally, the semipartial R^2 was calculated for each fixed effect by using the "nsj" method from the r2glmm package [17].

2.3.2. Performance Development per Subject. Since the PT also publishes results broken down by medical discipline or subject, all questions included in the test are routinely classified according to a list of 27 predetermined subjects; we used this classification to compute the absolute variations in the percentage share of correct answers for every subject-specific question subset. These question subsets are the same for both tests in each PT pair; therefore, a direct comparison between them is methodologically sound.

3. Results

3.1. Participants. The final datasets consisted of 13,372 tests for pair PT36-PT41, 13,121 for pair PT37-PT42, and 13,822 for pair PT38-PT43. Only serious test takers from nine

faculties were kept (see Figure 1 for the flow chart of data selection).

3.2. Overall Test Performance. The most recent test of each PT pair showed a higher mean score compared to the previous test in the same pair, with this difference becoming more pronounced over time (PT pair 36–41 : 2.53 (95% CI: 1.31–3.75), PT pair 37–42 : 3.72 (2.57–4.88), and PT pair 38–43 : 5.66 (4.63–6.69) (see Figure 2 and Supplementary Material Tables 1–3); this could signify a sustained knowledge increase throughout the whole period examined.

The variable "semester" was found to be the most influential fixed effect regarding student performance (>4.3 difference in mean between tests for every PT pair), implying that the mean score for each semester increases on average at least 4.3 points with respect to the previous semester regardless of other factors. This result is in line with expectations and reflects the usual knowledge increase of participants as they advance towards the completion of their degrees.

Regarding the interaction between test and semester, the results for each PT pair do not show a uniform picture. The values obtained for PT pairs 36–41 and 38–43 are −0.32 and −0.3; these negative figures imply that the growth of mean scores is stronger in earlier semesters and dwindles somewhat for more advanced students. This is not true in the case of PT pair 37–42, where the corresponding value (0.04) indicates that mean scores increase evenly throughout all semesters.

According to the intraclass correlation of all three models, university-related random effects do not generally add much variance to the obtained scores (PT pair 36–41 : 0.14, PT pair 37–42 : 0.06, and PT pair 38–42 : 0.04), which means that test results from the same university show comparatively low levels of within-cluster correlation. For all three models, the conditional R^2 lie around 0.48 and 0.56, respectively (i.e., the models explain 48% to 56% of variances in test scores). This is an expected outcome since variations in individual performance between participants belonging to the same university and semester are not covered by any model parameter; here, we have preferred to explore the evolution of test results for whole student cohorts instead of focusing on performance imbalances between students of the same group. From a methodological point of view, it is also worth mentioning that we have intended to model a numerical variable using a mixture of numerical and categorical variables, which may also have a negative effect on the value of R^2 (for the complete model results, see supplementary material Tables 1–3; the distribution of correctly answered questions of each PT pair per semester can be seen in Figure 3).

3.3. Development of Performance per Subject. As can be noted in Figure 4, subjects such as epidemiology, anesthesiology, or gynecology stand out markedly above the rest in terms of performance, while others (e.g., Urology, Dermatology,
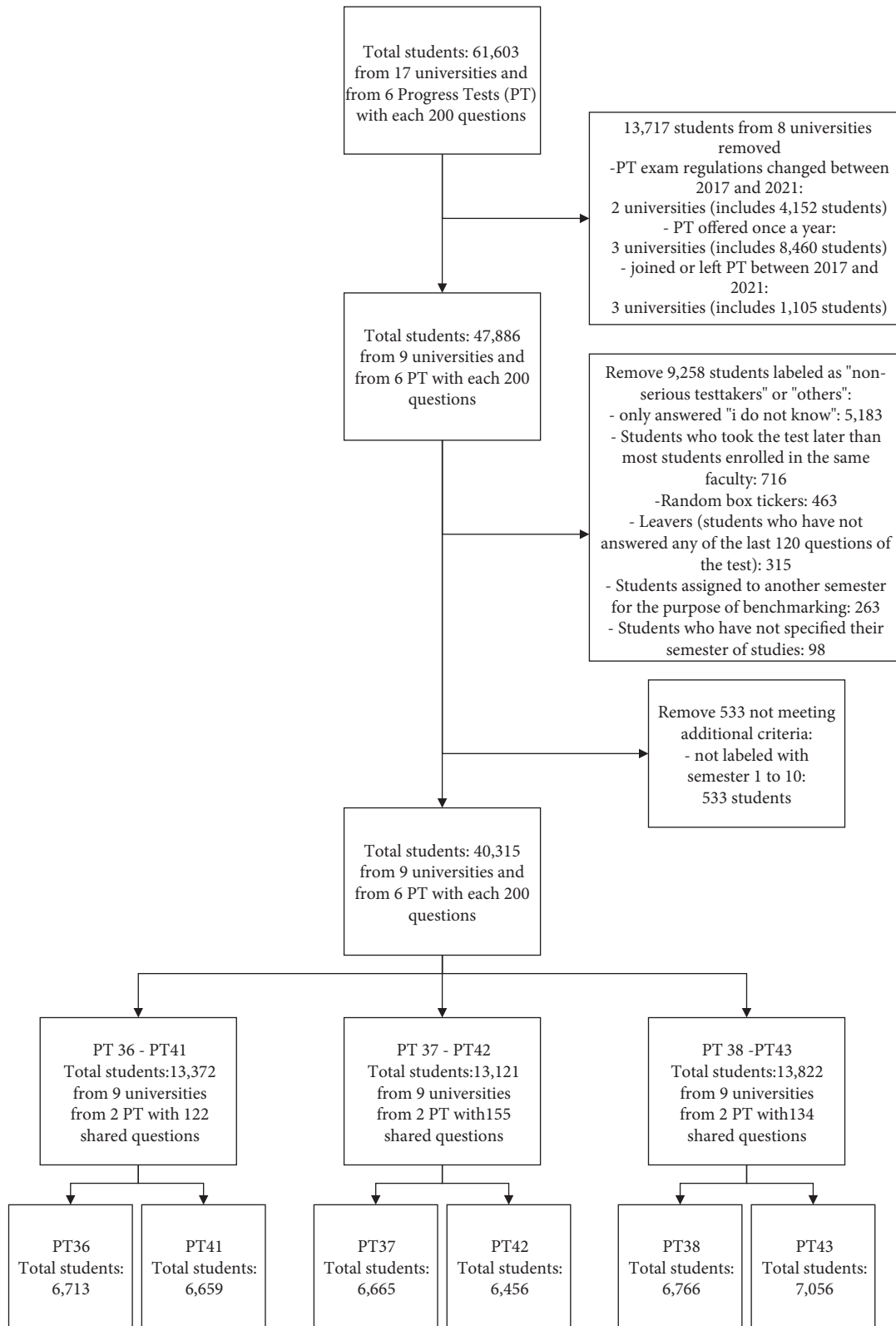
Figure 1: Flowchart of data cleaning process and allocation into different PT pairs.

Ophthalmology, or Otorhinolaryngology) show stagnant results. Detailed figures are included in the supplementary material (Table 4).

The medical discipline with the most noteworthy evolution is epidemiology (epi), whose share of correct answers in PT43 increased by 22.56 percentage points with respect to
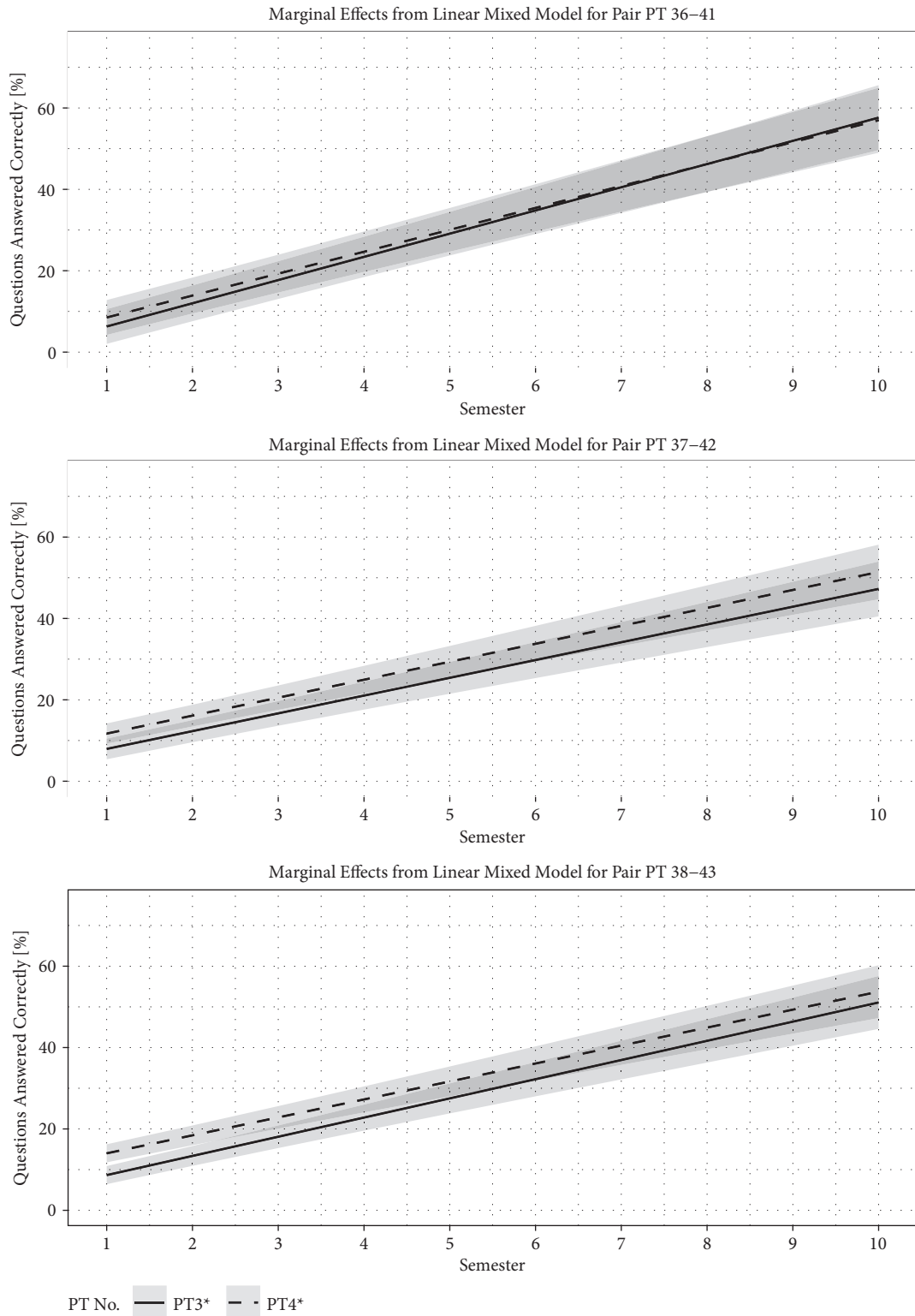
FIGURE 2: Results of the linear mixed model described in the article for each of the three PT pairs considered. Every PT shares a considerable number of questions with the PT that took place five semesters before; tests were conducted at nine different German and Austrian faculties. In these graphics, test modalities are fixed by their weighted estimates. The x-axis represents the semesters, while the achieved relative scores are shown on the y-axis. The dashed line represents the most recent test in each pair (41 = winter semester 2019/20, 42 = summer semester 2020, and 43 = winter semester 2020/21); the gray areas are the respective 95% confidence intervals. The difference in mean scores for students in earlier semesters (36 = summer semester 2017; 37 = winter semester 2017/18; 38 = summer semester 2018) increases with every successive pair, while the results of the pair PT37-PT42 seem to equally improve throughout all semesters; the scores of the other two pairs converge with increasing semesters. PT, Progress Test; ePT, electronic Progress Test.
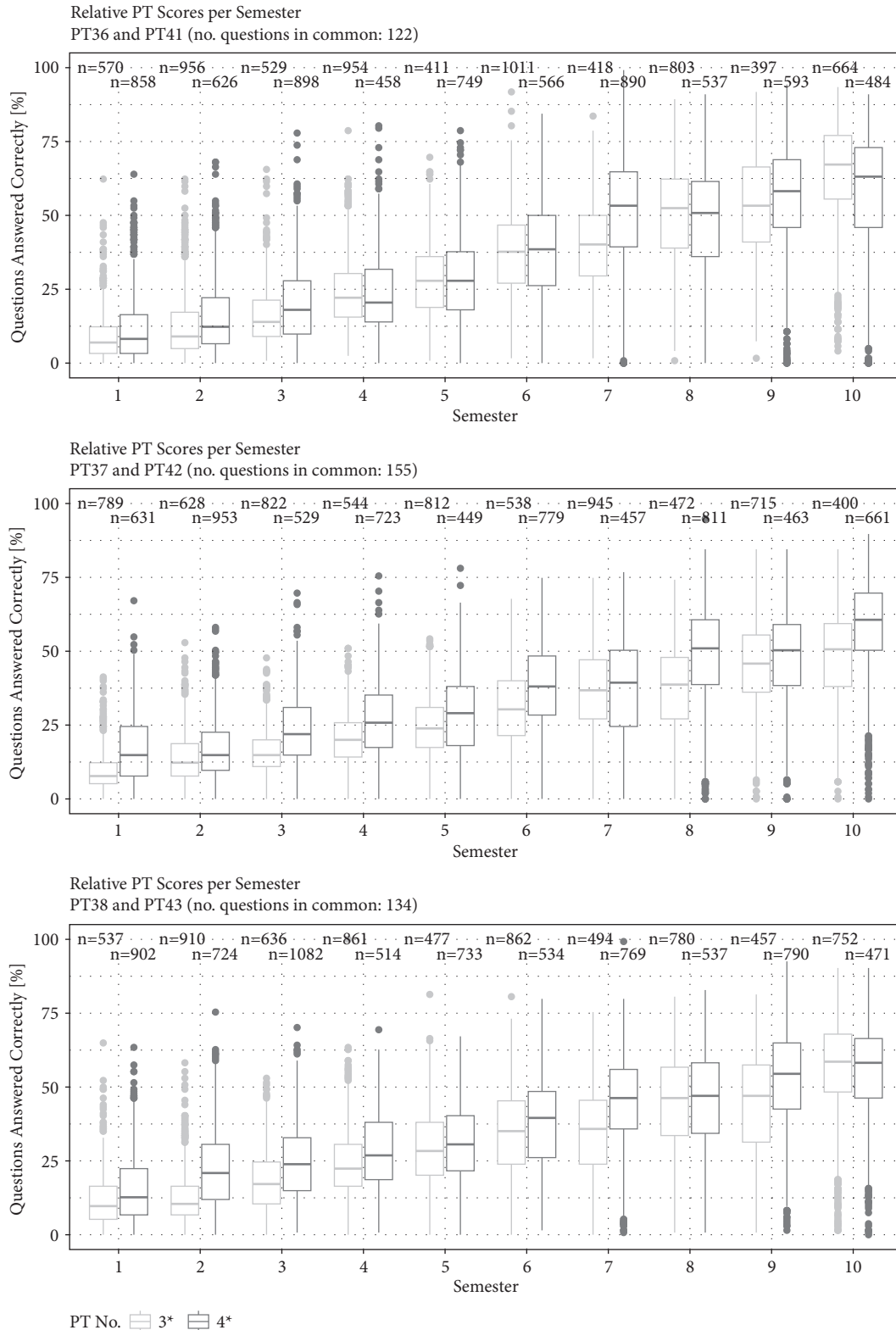
FIGURE 3: Relative PT scores per semester. Each PT shares a considerable number of questions with the PT that took place five semesters prior; therefore, PTs conducted five semesters apart from each other can be naturally paired together (PT pair). Boxplots show an overview of the percentage share of correctly answered questions in each PT pair by students enrolled at nine faculties in Germany and Austria. Data are disaggregated by semester; plots in light gray refer to data from earlier PTs (36 = summer semester 2017, 37 = winter semester 2017/18, 38 = summer semester 2018), while plots in dark gray represent later PTs (41 = winter semester 2019/20, 42 = summer semester 2020, and 43 = winter semester 2020/21, respectively). For each semester and PT, the number of participants is indicated above the corresponding boxplot. PT, Progress Test.
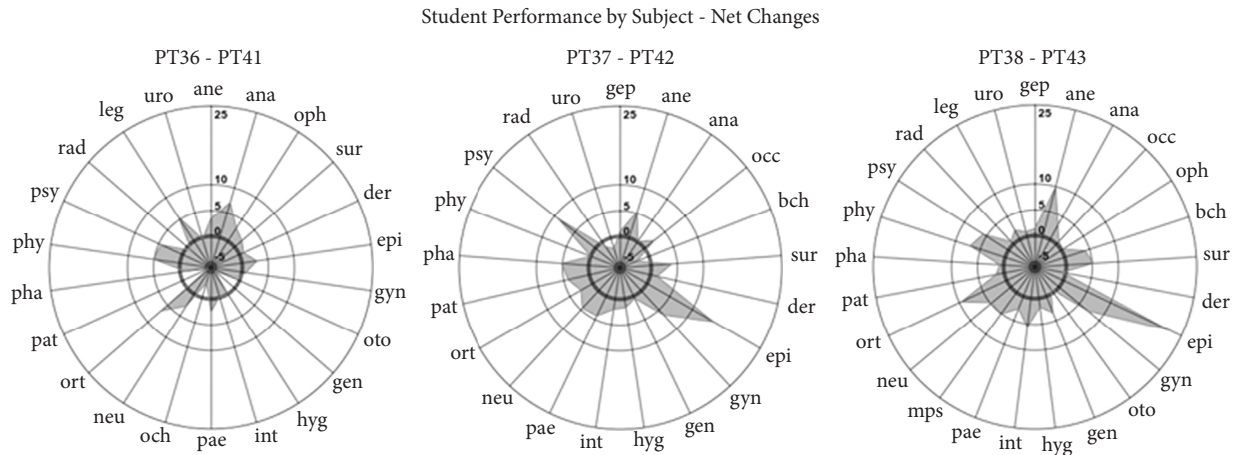
FIGURE 4: Radar charts showing variations in the percentage share of correct answers within each PT pair disaggregated by subject. Every PT shares a considerable number of questions with the PT that took place five semesters before. Data for these charts were collected from tests that took place from April 2017 to December 2020 (36 = summer semester 2017, 37 = winter semester 2017/18, 38 = summer semester 2018; 41 = winter semester 2019/20, 42 = summer semester 2020; 43 = winter semester 2020/21); further data treatment was performed in March 2021. Axis labels refer to absolute changes in percentage shares, with the thicker black circumference corresponding to 0% (no change). For every PT pair, only medical disciplines with at least two associated questions are featured in the corresponding graph. PT, Progress Test.

that of PT38; this is to be compared with an all-subject average gain of 4.21 percentage points in the same examination cycle. As a further reference point, the percentage of correct answers in epidemiology-related questions increased only by 2.92 points in PT41 in comparison with PT36; by the next examination cycle (PT42/PT37), the performance increase for epidemiology-related questions reached 14.76 percentage points, positioning itself clearly as the better performing subject while it had been ranked 8th in the previous cycle.

4. Discussion

COVID-19 pandemic lockdowns triggered sweeping changes in virtually all areas of society. Medical education was no exception to this rule: most faculties switched to online teaching, either reducing practical lectures and patient contact or even cancelling them altogether. These changes took place against a backdrop of fear and concern about various aspects of medical teaching and learning [18]. We used PT data to investigate the impact of these events on knowledge gain.

According to our analysis, both tests conducted during the pandemic (PT42 in April 2020 and PT43 in November 2020) show a relevant increase in mean scores of 3.72 and 5.66, respectively, when compared to previous tests belonging to the same pair (PT37 and PT38, respectively). With a mean score of 2.53, this effect is not so strong in the case of the PTs that took place before the pandemic (PT41 and PT36). This is mirrored by the net changes per subject; while the prepandemic pair shows an average linear increase of 1.40%, this effect is much stronger for pairs PT37-PT42 (3.05%) and PT38-PT43 (4.21%). There are a few medical disciplines that even emerge as winners from the current situation; in fact, the outstanding performance improvement in epidemiology-related questions might be understood as a side effect of the pandemic.

There are a wide variety of circumstances that might influence academic performance on both individual and collective levels. However, the PT examination framework remained almost unchanged between PT36 and PT43 save for the implementation of technologically advanced test modalities at some of the participating medical schools. These test modalities were thus included in our models as fixed effects to quantify and delimit their influence on test scores. One effect we could not directly account for was COVID-19, which induced major study environment changes in the participating faculties in Germany and Austria over the period between PT41 and PT43. We, therefore, link the majority of the unaccounted differences in performance results to these changes.

The results reported in the literature are not unanimous; some findings describe negative trends [3] or impacts [4, 19] of the pandemic on the academic performance of medical students. On the other hand, there are findings where medical students have stressed the benefits of online lectures [6] and also studies concluding that their cognitive performance remained the same [5] or improved [20].

Our results suggest a sustained performance improvement widely spread across comparison groups and participating faculties; this improvement is also noticeable at the subject level, although its distribution among medical disciplines is somewhat uneven. In this context, one must keep in mind that the PT is a formative test that assesses the end objectives of the curriculum in contrast to summative examinations of specific course content [21].

It must be mentioned that first-semester students also performed better in the tests conducted during the pandemic; these students wrote their PT only a few weeks into their studies. This outcome may have been partly driven by new admission criteria introduced by some of the participating medical schools in 2020 [22]; nevertheless, we recommend further research at this point.

In summary, there is good evidence that the shift to distance learning prompted by COVID-19 resulted in an increased knowledge gain. However, we must remark that our study was limited to the domain of theoretical knowledge; further research on how the pandemic conditions may have also affected the acquisition of practical skills would be much needed in order to build a complete view of the broader topic.

4.1. Limitations

4.1.1. Scope of the Study. One must keep in mind that the percentages between the different test pairs are not comparable without conditions as the difficulties of the chosen questions may differ.

On another note, this study only covers possible short-term effects of the COVID-19 pandemic on medical students; an investigation of medium-term or even long-term effects would require more prolonged monitoring of results.

4.1.2. Regression. We treated overall score changes between semesters as if they were linear, but there might be certain semesters where the extent of the knowledge increase differs from the average. Variances in the model were heterogeneous, which might lead to underestimated standard errors.

5. Conclusion

The shift to distance learning prompted by COVID-19 resulted in an increased knowledge gain compared to Progress Tests administered before the pandemic.

These findings could also be relevant in the future since they are descriptive (at least to some extent) of how medical schools in Germany and Austria used digitalization and online learning as tools to cope with the impact of an unforeseen critical event with major consequences. As such, these adjustments and their effects should not be overlooked since they could serve as a "dress rehearsal" [23] for future challenges on a global level. It is important to keep in mind that we are not able to give forecasts regarding effects on the practical skills or mental state of students.

One way or another, the current worldwide push for digital education makes it appropriate to build a corpus of evidence on its effects on student experience, even if now we are only able to discuss short-term developments.

Abbreviations

COVID-19: Coronavirus disease 2019
PT: Progress Test
CI: Confidence Interval
ePT: Electronic Progress Test.

Data Availability

The datasets generated during and/or analyzed during the current study are not publicly available for data security reasons but are available from the corresponding author on reasonable request and after approval of the Progress Test cooperation partners and an extended ethical approval.

Ethical Approval

Ethical approval was granted by the Ethics Committee at the Charité (EA4/242/20). All methods were performed according to relevant guidelines and regulations. Regarding the usage of data about student performance in Progress Tests, the authors also refer to the local university law (BerlHG; §6) and the local examination regulations.

Disclosure

The preprint and its history can be accessed at https://www.researchsquare.com/article/rs-786483/v2.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

MM and VS outlined the concept and design of the article. IRA and MS developed the study design and prepared and analyzed the data. All authors discussed and evaluated the results. VS and JS drafted the introduction section. MM, IRA, and MS drafted the methods and results. VS and MM drafted the discussion and conclusion section. All authors contributed to the manuscripts' revision. IRA performed the manuscripts' language editing. All authors read and approved the final manuscript.

Acknowledgments

Supplementary Materials

The supplementary material ("Supplementary Material.pdf") contains four tables, namely, respective mixed regression model outputs for pairs PT36-PT41, PT37-PT42, and PT38-PT43 (Supplementary Tables 1–3); detailed figures on subject development are given in Supplementary Table 4. (*Supplementary Materials*)

References

[1] A. M. Rogowska, I. Pavlova, C. Kuśnierz, D. Ochnik, I. Bodnar, and P. Petrytsa, "Does physical activity matter for the mental health of university students during the COVID-19 pandemic?" *Journal of Clinical Medicine*, vol. 9, no. 11, p. 3494, 2020.

[2] A. Vahedian-Azimi, M. Moayed, F. Rahimibashar, S. Shojaei, S. Ashtari, and M. Pourhoseingholi, "Comparison of the severity of psychological distress among four groups of an Iranian population regarding COVID-19 pandemic," *BMC Psychiatry [revista en Internet]*, vol. 20, 2020.

[3] N. Al-Husban, A. Alkhayat, M. Aljweesri et al., "Effects of COVID-19 pandemic on medical students in Jordanian universities: a multi-center cross-sectional study," *Annals of Medicine and Surgery*, vol. 67, Article ID 102466, 2021.

[4] A. Harky, D. Karimaghaei, H. Katmeh, and S. Hewage, "The impact of COVID-19 on medical examinations," *Acta Biomed Atenei Parm*, vol. 91, no. 4, 2020.

[5] N. Khoueiry Zgheib, A. Ali, and R. Sabra, "Experience with forced transition to online learning during the COVID-19 pandemic: students' cognitive performance and their perceptions of teaching," *The Asia Pacific Scholar*, vol. 6, no. 3, pp. 45–55, 2021.

[6] J. W. Kim, S. J. Myung, H. B. Yoon, S. H. Moon, H. Ryu, and J.-J. Yim, "How medical education survives and evolves during COVID-19: our experience and future direction," *PLoS One*, vol. 15, no. 12, Article ID e0243958, 2020.

[7] K. Osterberg, S. Kölbel, and K. Brauns, "Der progress test Medizin: erfahrungen an der Charité Berlin," *GMS Zeitschrift für Medizinische Ausbildung*, vol. 23, no. 3, 2006.

[8] J. E. Kämmer, W. E. Hautz, and M. März, "Self-monitoring accuracy does not increase throughout undergraduate medical education," *Medical Education*, vol. 54, no. 4, pp. 320–327, 2020.

[9] Y. Karay, S. K. Schauber, C. Stosch, and K. Schuettpelz-Brauns, "Can computer-based assessment enhance the acceptance of formative multiple choice exams? A utility analysis," *Medical Teacher*, vol. 34, no. 4, pp. 292–296, 2012.

[10] K. Schüttpelz-Brauns, M. Hecht, K. Hardt, Y. Karay, M. Zupanic, and J. E. Kammer, "Institutional strategies related to test - taking behavior in low stakes assessment," *Advances in Health Sciences Education*, vol. 25, no. 2, pp. 321–335, 2019.

[11] K. Schüttpelz-Brauns, M. Kadmon, C. Kiessling, Y. Karay, M. Gestmann, and J. E. Kämmer, "Identifying low test-taking effort during low-stakes tests with the new Test-taking Effort Short Scale (TESS) – development and psychometrics," *BMC Medical Education*, vol. 18, pp. 101–110, 2018.

[12] Y. Karay, S. K. Schauber, C. Stosch, and K. Schüttpelz-Brauns, "Computer versus paper—does it make any difference in test performance?" *Teaching and Learning in Medicine*, vol. 27, no. 1, pp. 57–62, 2015.

[13] M. G. Simkin and A. McLeod, "Why do college students cheat?" *Journal of Business Ethics*, vol. 94, no. 3, pp. 441–453, 2010.

[14] E. M. Anderman and A. C. Koenka, "The relation between academic motivation and cheating," *Theory Into Practice*, vol. 56, no. 2, pp. 95–102, 2017.

[15] Z. M. Nouns and W. Georg, "Progress testing in German speaking countries," *Medical Teacher*, vol. 32, no. 6, pp. 467–470, 2010.

[16] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, 2015.

[17] B. C. Jaeger, "r2glmm: computes R squared for mixed (multilevel) models," 2017, https://cran.r-project.org/package=r2glmm.

[18] M. N. Ferrel and J. J. Ryan, "The impact of COVID-19 on medical education," *Cureus*, vol. 12, no. 3, p. e7492, 2020.

[19] S. Andersen, G. Leon, D. Patel, C. Lee, and E. Simanton, "The impact of COVID-19 on academic performance and personal experience among first-year medical students," *Medical Science Educator*, vol. 32, no. 2, pp. 389–397, 2022.

[20] M. Sekine, M. Watanabe, S. Nojiri et al., "Effects of COVID-19 on Japanese medical students' knowledge and attitudes toward e-learning in relation to performance on achievement tests," *PLoS One*, P. Mittal, Ed., vol. 17, no. 3, Article ID e0265356, 2022.

[21] A. Freeman, C. Van Der Vleuten, Z. Nouns, and C. Ricketts, "Progress testing internationally," *Medical Teacher*, vol. 32, no. 6, pp. 451–455, 2010.

[22] KMK, "Studienplatzvergabe im zentralen vergabeverfahren: kultusministerkonferenz verabschiedet entwurf des staatsvertrags über die hochschulzulassung," 2018, https://www.kmk.org/aktuelles/artikelansicht/studienplatzvergabe-im-zentralen-vergabeverfahren-kultusministerkonferenz-verabschiedet-entwurf-des.html.

[23] C. Nikendei, A. Cranz, and T. J. Bugaj, "Medical education and the COVID-19 pandemic - a dress rehearsal for the "climate pandemic"?" *GMS Journal for Medical Education*, vol. 38, no. 1, 2021.