

# Short Term Effect of the COVID-19 Pandemic On Progress Test Outcomes and Emotional State of Medical Students: A Quantitative and Qualitative Analysis

## Progress Test Medizin Charité Universitätsmedizin Berlin

Charité Universitätsmedizin Berlin Corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin

## Victoria Sehy (✉ [victoria.sehy@charite.de](mailto:victoria.sehy@charite.de))

Charité Universitätsmedizin Berlin Corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin

## Iván Roselló Atanet

Charité Universitätsmedizin Berlin Corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin

## Miriam Sieg

Charité Universitätsmedizin Berlin Corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin

## Jana Struzena

Charité Universitätsmedizin Berlin Corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin

## Maren März

Charité Universitätsmedizin Berlin Corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin

---

## Research Article

**Keywords:** Progress Test, academic performance, COVID-19, wellbeing, distance learning, NLP, LDA, learning behaviors, undergraduate medical education

**Posted Date:** September 21st, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-786483/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

# Abstract

**Background** The COVID-19 pandemic has been the source of many challenges for medical students worldwide. The authors examined short-term effects on knowledge gain as well as shifts in learning behavior and study-related emotional states.

**Method** The development of knowledge gain was measured comparing the outcomes of shared questions within Progress Test (PT) pairs. The authors used mixed-effect regression models and compared the absolute variations in the percentage of correct answers per subject. Three successive test pairs were analyzed in this manner: 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).

A survey including closed and open-ended questions was also carried out in January 2021 with the purpose of assessing the learning behavior and emotional state of participants. Open-ended responses were analyzed using Latent Dirichlet Allocation.

**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), PT38-PT43: 5.59 (4.37-6.81)). Analogously, an increase in the share of correct answers was observed for most medical disciplines, with Epidemiology showing the most remarkable upsurge.

N=2,715 students from eleven different German-speaking faculties participated in the survey. Respondents were mostly positive towards online lectures, which were perceived as clearly beneficial, allowing for more time and flexibility. On the other hand, the suspension of practical lessons and alleged communicational and organizational shortcomings were seen as the main disadvantages.

28% of the students did not perceive negative impacts on their emotional state regarding their studies, however, 20% of the surveyed students found it difficult to cope with the lack of social contacts, with an additional 8% of them claiming to feel lonely, demotivated or abandoned.

**Conclusion:** Overall, PT performance improved during the pandemic. Students see advantages in online lectures, but disadvantages in the cancellation of practical lectures; they miss their former social interactions and some even show signs of emotional distress.

## Background

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 include usually an extensive practical component implying regular contact to 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 emotional state of students has been subject to research. Mahdy, 2020 found that students do value some advantages of online learning, finding it more flexible and convenient than face-to-face tuition; they also point out that online learning saves time and provides opportunities for self-study. According to this source, this is sometimes countered by the lack of availability of internet and/or computer equipment, as well as by difficulties in adequately organizing practical lessons and insufficient levels of communication and interaction leading to negative effects on study concentration; all in all, 97% of Mahdy's respondents stated that the COVID-19 pandemic lockdown has affected their academic performance (3). Respondents of Meo et al, 2020 confirmed these findings; they also alluded to a decrease in their work performance and study time, plus a deterioration of their psychological well-being during quarantine. Meo et al. found that many students felt an emotional workload; 38% of the students they surveyed claimed to feel hopeless, exhausted, and emotionally unresponsive (4).

In general, an interaction between environmental and emotional characteristics and academic performance has been proven to exist (5), however, there is still insufficient data available to quantify the impact of the pandemic on knowledge acquisition. We therefore used data from Progress Tests (PTs) in medical education to investigate short-term effects on knowledge gain; we also conducted a survey of medical students enrolled at the participating faculties with the purpose to find out to what extent the pandemic has impacted student life.

The PT is a formative test of 200 multiple choice questions at graduate level, which provides feedback to students on knowledge and knowledge gain during their course of study (6). In Germany and Austria, it is provided to about 11,000 students of 14 faculties around the beginning of each semester. The “winter semester PT” takes place usually from October to December, while the “summer semester PT” is conducted from April to June.

In summary, we intend to address the following three main questions throughout this study using data from PT as well as survey responses given by participating students on a voluntary basis:

- 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?
- Are there differences at specialty level? A relevant question here is whether the observed changes were similar for all fields of study or there were remarkable differences in performance depending on the medical discipline considered.
- How did medical students react to changes in their study environment due to the pandemic?

## Materials And Methods

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” respectively “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 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 had to face 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”.

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. Firstly, their PT exam regulations should not have changed between the summer semester of 2017 and the winter semester of 2020 (e.g., from mandatory to voluntary test-participation), and secondly, they should offer the test each semester as opposed to once in a year (even if students are not required to take it every semester).

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 the students to each question as well as the semester to which they belong, the faculty where they study and whether their participation in the test is considered as ‘serious’ or not. ‘Non-serious’ participants are excluded from the calculation of comparison groups since the validity of results would be jeopardized otherwise (8). ‘Non-serious’ is presumed when the number of answered questions is below a previously established threshold or answers show a pattern of randomness (8, 9).

Since both tests belonging to the same pair share most of their questions, students who took both tests in a pair were confronted with the shared questions twice, while the rest were shown 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 was set to 0.01.

## Regression Models

For each PT pair, we fitted a linear mixed effect model with random intercept and slope and with the overall PT score as the outcome variable. Four fixed effects predictors were set: test number, semester of study, interaction between test number and semester of study, and test modality (electronic PT (ePT) vs. non-ePT). The medical school where each test was administered (random intercept) and the interaction between medical school and semester of study (random slope) were chosen as random effect predictors. This choice assumed (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 and the lmer function from the R-package lme4 (10) for fitting the models, additionally, the semi-partial  $R^2$  was calculated for each fixed effect by using the “nsj” method from the r2glmm package (11).

## Subject Development

All questions included in PT are routinely classified according to the medical discipline or subject to which they are closest; a list of 27 predetermined subjects is used for that purpose. In order to determine how our student performance data broke down by subject, we computed the absolute variations in the percentage share of correct answers for the subsets of questions generated along these lines (these questions are the same for both tests in each PT pair; therefore, a direct comparison between them is methodologically sound). For every PT pair, only medical disciplines with at least two associated questions are featured in the corresponding graph.

## Survey

The survey was based on the work of Mahdy and Schaubert et al. (3, 5) and conducted using the software SoSci Survey (12). Participation was voluntary and anonymous; medical students enrolled at participating faculties were sent an e-mail with a link to the survey website. The questionnaire included three closed questions regarding the faculty and semester the student belongs to as well as their level of satisfaction with their own academic performance (i.e., summative assessments). It continued with five open-ended questions about emotional and environmental aspects (i.e., learning strategies; advantages and disadvantages of new study environments, and study-related emotional states in the context of the COVID-19 pandemic; see Additional file 1 for the complete questionnaire).

Open-ended questions are considered more difficult to analyze than their closed counterparts, as human coding is almost always used (7). Topic modelling is increasingly performed with the aid of machine learning algorithms; in its simplest form, this is an automated, unsupervised process (13) in which each document is supposed to contain a certain number of themes or topics. These topics are derived by identifying the words that are statistically closest to each other (14).

We used the Latent Dirichlet Allocation (LDA) algorithm, which assigns both words and documents (in the sense of ‘sequences of words’) to one or more topics. This algorithm is based on the idea that documents are represented as random mixtures over latent topics, where each topic is characterized by a probability distribution over words. The number of words among which to choose as well as the number of possible topics per document must be determined in advance; topics are then allocated to their corresponding words following a probability distribution conditioned on the topic. The two fundamental parameters of the LDA algorithm are known as  $\alpha$  and  $\beta$ ;  $\alpha$  defines a Dirichlet probability distribution according to which topics are selected, while  $\beta$  determines the selection probabilities of each possible word given each possible topic (15). This approach requires some preparation of the data. We converted all words into lowercase and lemmatized them

using spaCy (16) Some, but not all stopwords were removed (14); additionally, lemmatization was carried out where necessary (e.g., „Bibliotheken“ (German for libraries) and its common abbreviation „BIB“).

We used CountVectorizer from sklearn to build feature vectors from the entries and applied LDA for topic modelling, using Gridsearch for model optimization (15, 17, 18).

It is known that topic modelling does not always deliver results that are easy to understand or correlate with human judgements(19). We therefore adjusted the number of topics, starting with the ideal number given by the algorithm until reaching a good fit from a human perspective, and provided afterwards a description for each topic.

## Statistical Analysis

Data were analyzed using IBM SPSS Statistics for Windows, version 27 (IBM Corp., Armonk, N.Y., USA), R for Windows, version 3.6.1 (R Core Team, Vienna, Austria) and Python, version 3.8 (Python Software Foundation, Beaverton, USA).

## Results

### General results

The final datasets consisted of 13,365 tests for pair PT36-PT41, 13,107 for pair PT37-PT42, and 13,808 for pair PT38-PT43. Only serious test takers from 9 faculties were kept; that is, a mean of 65.6% of tests included in the initial dataset remained in our study after non-serious test takers and students from non-participating universities were removed from the datasets (see Additional file 2 for the exact figures).

An overview of the correctly answered questions of each PT pair per semester can be seen in Fig. 1.

Furthermore, students who were shown the shared questions twice did not outperform students who were shown these questions only once (t-statistic - 0.71, p-value 0.87 and effect size (Cohen's d) -0.01 for pair 36-41; 1.97, 0.05, 0.07 for pair 37-42, and 2.43, 0.02, 0.09 for pair 38-43, respectively) (Additional file 3).

### Regression

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 (pair PT36-PT41: 2.53 (95% CI: 1.31-3.75), pair PT37-PT42: 3.72 (2.57-4.88), pair PT38-PT43: 5.59 (4.37-6.81), Fig. 2).

However, the semi-partial  $R^2$  for these variables never exceeded 0.01. As expected, the variable "semester" was proven to be the most influential fixed effect regarding student performance (> 4.3 difference in mean between tests for every PT pair). As for the interaction between test and semester, the results for each pair do not show a uniform picture; in pairs PT36-PT41 and PT38-PT43 we see that the growth of mean scores is stronger in earlier semesters and dwindles somewhat for more advanced students, while this is not true in the case of pair PT37-PT42, where 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 (pair PT36-PT41: 0.14, pair PT37- PT42: 0.06, pair PT38-PT43: 0.04), which means that test results from the same university show comparatively low levels of within-cluster correlation. For all three models, the marginal and conditional  $R^2$  lie around 0.48 and 0.54, respectively (see Additional file 4 to 6 for complete mixed regression model output).

## Subject Development

As can be noted in Fig. 3, some subjects stand out markedly above the rest in terms of performance, while others show stagnant results.

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 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. For more detailed figures on subject development see also Additional file 7.

## Survey

A total of 2,715 students from eleven different universities took part (two faculties, that were not included in the quantitative study participated in the survey). The number of respondents per semester ranged from N = 34 to N = 541 with the 3rd semester being the one with the most respondents.

To evaluate satisfaction with their academic performance during the pandemic, participants were asked to choose among 5 smileys ranging from “very sad” = 1 to “very happy” = 5 (Mean = 3.7; SD = 1.2). As shown in Fig. 4, students across all semesters and faculties were quite satisfied with their performance in the summative examinations. Students’ open-ended responses contained a wealth of information, touching many topics and bringing up disadvantages, advantages, and emotions in any given answer, regardless of what was asked. The LDA model allows each document to exhibit multiple topics. The result is a distribution of probabilities for each document belonging to each topic, the greatest probability being the resulting ‘dominant topic’(15). This means that the allocation of several answers to a certain topic does not necessarily imply that the people who gave these answers do not agree to some extent with any of the other possible topics.

2,265 students replied to the question regarding adaptations of the learning environment. According to the LDA, 25% stated that the biggest change for them was having to study at home, mostly alone. 22% invested a greater amount of time in their studies, with varying degrees of success. 17% claimed to have improved their time management as they were no longer required to travel to campus or attend lectures. 20% focused on structuring their learning and everyday life. 16% did not change their learning behavior at all or changed it only slightly; this group includes first semester students.

2,306 students commented on advantages. 80% referred to online lectures, highlighting three main benefits: more time available due to cancelled lectures (31%), learning at their own pace (25%) and flexible time management (24%). 20% did not perceive any positive change.

From the application of the LDA to the responses of the 2,333 students citing disadvantages we estimate that 25% missed social contacts such as fellow students; 21% missed their familiar learning environment (library, study groups), and 17% considered the lack of practical lessons and the limited choice of courses available to be a major disadvantage. Some students also felt abandoned by their faculty (19%); others claimed that the situation was uncertain and said to be lacking adequate communication from their faculties (18%).

2,462 students commented on their emotional state regarding their studies. The LDA analysis suggests that 28% of them suffered no study-related emotional distress; on the contrary, a further 24% reported changes in motivation, with most of them experiencing a decrease in this regard. 20% of students hinted at organizational flaws, while another 20% felt lonely, missed the exchange with fellow students or experienced anxiety about their exams. 8% of answers might be interpreted as expressing feelings of distress or even depression.

## Discussion

COVID-19 pandemic lockdowns had 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. We investigated the impact of these events on knowledge gain using PT data and analyzed personal perceptions regarding changes in their learning environment by performing a survey among students of the PT consortium.

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 5.59 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; whilst the pre-pandemic 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 which 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.

Results on knowledge gain correlate with high satisfaction levels with summative assessment results during both semesters (mean of 3.7 on a 1–5 scale).

There are a vast variety of circumstances which 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 were therefore included in our models as random effects to quantify and delimit their influence on test scores. Other than that, we reckon that COVID-19 was the only issue known to have induced major study environment changes in the participating faculties in Germany and Austria within the period between PT36 and PT43.

The findings from our survey are mostly in line with existing literature (3, 4). When asked about teaching-related changes implemented in the wake of the pandemic, an overwhelming majority of students (80% of 2,306) claimed to perceive at least some advantages in online lectures; as such, they highlighted time saving, flexibility and the possibility of taking more active roles in their learning process. On the other hand, students place a high value on practical tuition and put emphasis on their need to be adequately prepared for the profession; therefore, the suspension of practical lessons was seen as a major disadvantage. Some students also hinted at inadequacies in both faculty organization and student communication. Despite this, almost one third of students did not report to have experienced any study-related emotional distress. Some, however, mentioned feelings of loneliness, loss of motivation and a sense of having been abandoned by their faculty.

## Limitations

### Scope of the study

This study only covers possible short-term effects of the COVID-19 pandemic on medical students; an investigation on medium-term or even long-term effects would require a more prolonged monitoring of results. We must also remark that our study was limited to the field of theoretical knowledge; further research on how the pandemic may have also affected the acquisition of practical skills would be much needed in order to build a complete view of the broader topic.

## Regression

We assumed that overall score changes between semesters are 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.

## Survey

It must be considered that some responses to open-ended questions touched many aspects; because of that, assigning these responses to only one main topic was not entirely straightforward. There were also instances where very similar sentences would express opposite meanings, thus making the training of the LDA somewhat challenging.

## Conclusion

The shift to distance learning prompted by COVID-19 resulted in an increased knowledge gain compared to Progress Tests administered before the pandemic; this is consistent with the overwhelming acceptance of online lectures, which many participants in the survey would like to keep even after the end of the pandemic. In contrast, students identified the cancellation of practical lessons as a source of insecurity regarding their studies; organizational shortcomings and lack of communication from faculties were also the subject of some criticism. While a solution for practical lessons might not be at hand quickly, there could still be some room for faculties to reassess their student communication policies and to introduce measures to help students in need, such as implementing mentoring programs or setting up online learning groups.

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

The implementation of an LDA-assisted survey in which students commented freely on their personal subjective experiences during these times provided us with very valuable insight about their outlook on advantages and disadvantages of online learning; in addition, performance data from tests taken both after March 2020 and shortly beforehand allowed us to observe and calibrate the impact of the pandemic on every single subject. We believe that the body of information collected for this article might be useful to make the experience of studying a medical degree smoother and more enjoyable for students, thereby helping raise both their motivation and performance levels.

## Abbreviations

COVID-19: Coronavirus disease 2019; PT: Progress Test; CI: Confidence Interval; ePT: electronic Progress Test; LDA: Latent Dirichlet Algorithm; sklearn: scikit-learn

## Declarations

### **Ethical Approval and *Consent for Publication***

Ethical approval was granted by the Ethics Committee at the Charité (EA4/242/20). All methods were performed according to relevant guidelines and regulations. Informed consent was obtained from participants in our survey as they were invited via e-mail. It was stressed in the e-mail and the invitation message that participation was anonymous and voluntary and that the results of the study would be presented in an aggregate form. Students could decline to fill out the questionnaire if they did not consent to these terms and conditions. Neither responders nor non-responders could be identified in any way. Regarding the usage of data about student performance in Progress Tests, we refer to the local university law (BerlHG; §6) and the local examination regulations.

### **Availability of data and materials**

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.

## Competing interests

The authors declare that they have no competing interests

## Funding

The authors received no funding

## Authors' contributions

MM & VS formed the concept and design of the article. VS, JS & MM developed and conducted the survey. IRA & MS developed the quantitative Datas' study design, prepared and analyzed the data. MM analyzed the qualitative data via LDA. All authors discussed and evaluated the results of both, qualitative and quantitative data. VS & JS drafted the introduction section. MM, IRA & MS drafted the methods and results. VS & 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

The authors would like to thank the Partners of the *Progress Test Medizin* consortium for their valuable advice regarding this study and their participation.

The Progress Test Team members would furthermore thank Susanne Werner for her contribution regarding the survey.

### *Collaborating author names:*

**Victoria Sehy**, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, Progress Test Medizin, Charitéplatz 1, 10117 Berlin, Germany

**Iván Roselló Atanet**, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, Progress Test Medizin, Charitéplatz 1, 10117 Berlin, Germany

**Miriam Sieg**, 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, 10117 Berlin, Germany

ORCID: <https://orcid.org/0000-0001-9724-3129>

**Jana Struzena**, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, Progress Test Medizin, Charitéplatz 1, 10117 Berlin, Germany

**Maren März**, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, Progress Test Medizin, Charitéplatz 1, 10117 Berlin, Germany, ORCID: <https://orcid.org/0000-0002-2661-5076>

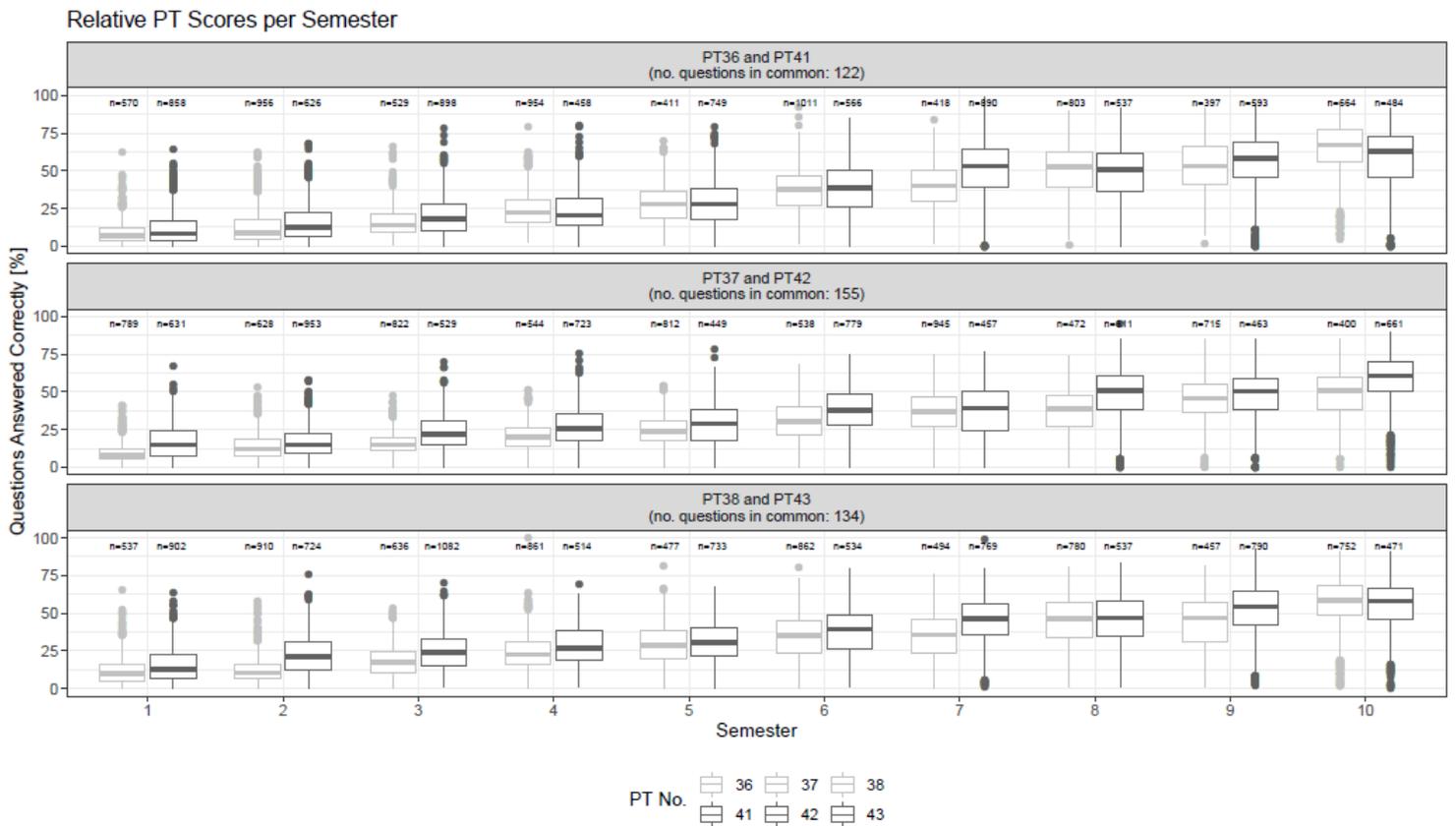
The complete results of our survey were published on the PT website ([https://progress-test-medin.charite.de/en/blog/survey\\_influence\\_of\\_the\\_pandemic\\_on\\_undergraduate\\_medical\\_studies\\_ai\\_supported\\_analysis/](https://progress-test-medin.charite.de/en/blog/survey_influence_of_the_pandemic_on_undergraduate_medical_studies_ai_supported_analysis/))

## References

1. Rogowska AM, Pavlova I, Kuśniercz C, Ochnik D, Bodnar I, Petrytsa P. Does Physical Activity Matter for the Mental Health of University Students during the COVID-19 Pandemic? *J Clin Med.* 2020;9(11):3494.

2. Vahedian-Azimi A, Moayed M, Rahimibashar F, Shojaei S, Ashtari S, Pourhoseingholi M. Comparison of the severity of psychological distress among four groups of an Iranian population regarding COVID-19 pandemic. *BMC Psychiatry* [revista en Internet] 2020 [acceso 30 de octubre de 2020]; 402(2020): 1–7. *BMC Psychiatry*. 2020;1–7.
3. Mahdy MAA. The Impact of COVID-19 Pandemic on the Academic Performance of Veterinary Medical Students. *Front Vet Sci*. 2020;7(October):1–8.
4. Meo SA, Abukhalaf DAA, Alomar AA, Sattar K, Klonoff DC. COVID-19 Pandemic: Impact of Quarantine on Medical Students' Mental Wellbeing and Learning Behaviors. *Pakistan J Med Sci*. 2020 May 11;36(COVID19-S4):43–8.
5. Schauber SK, Hecht M, Nouns ZM, Kuhlmeier A, Dettmer S. The role of environmental and individual characteristics in the development of student achievement: a comparison between a traditional and a problem-based-learning curriculum. *Adv Heal Sci Educ*. 2015;20(4):1033–52.
6. Osterberg K, Kölbl S, Brauns K. Der Progress Test Medizin: Erfahrungen an der Charité Berlin. *GMS Z Med Ausbild*. 2006;23(3)::Doc46.
7. Roberts ME, Stewart BM, Tingley D, Lucas C, Leder-Luis J, Gadarian SK, et al. Structural topic models for open-ended survey responses. *Am J Pol Sci*. 2014;58(4):1064–82.
8. Schüttpelz-Brauns K, Kadmon M, Kiessling C, Karay Y, Gestmann M, Kämmer JE. Identifying low test-taking effort during low-stakes tests with the new Test-taking Effort Short Scale (TESS) – development and psychometrics. *BMC Med Educ*. 2018;1–10.
9. Schüttpelz-Brauns K, Hecht M, Hardt K, Karay Y. Institutional strategies related to test – taking behavior in low stakes assessment. *Adv Heal Sci Educ*. 2019;(0123456789).
10. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. *J Stat Softw*. 2015;67(1).
11. Jaeger BC. Package 'r2glmm.' Packag 'r2glmm.' 2017;12.
12. Leiner, D. J. (2021). SoSci Survey (Version 3.2.23) [Computer software]. Available at <https://www.sosicisurvey.de>.
13. Ghahramani Z. Unsupervised Learning. In: Bousquet O, von Luxburg U, Rätsch G, editors. Berlin, Heidelberg: Springer Berlin Heidelberg; 2004. p. 72–112.
14. Campbell JC, Hindle A, Stroulia E. LATENT DIRICHLET ALLOCATION: EXTRACTING TOPICS FROM SOFTWARE ENGINEERING DATA. *The Art and Science of Analyzing Software Data*. Elsevier Inc.; 2015. 139–159 p.
15. Blei DM, Ng AY, Jordan MI. Latent Dirichlet Allocation. *J Mach Learn Res*. 2003;3:139–59.
16. spaCy [Internet]. Available from: <https://spacy.io/>
17. Sk. 1.4. Support Vector Machines [Internet]. [cited 2020 Sep 2]. Available from: <https://scikit-learn.org/stable/modules/svm.html>
18. Buitinck L, Louppe G, Blondel M, Pedregosa F, Mueller A, Grisel O, et al. API design for machine learning software: experiences from the scikit-learn project. 2013;1–15.
19. Hu Y, Boyd-Graber J, Satinoff B. Interactive topic modeling. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics,. 2011. p. 248–57.
20. Natural Language Toolkit [Internet]. Available from: <https://www.nltk.org/>
21. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. *J Mach Learn Res*. 2011;12:2825–30.

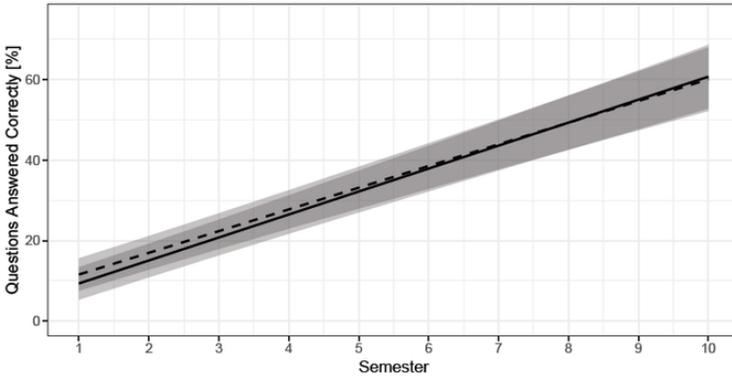
## Figures



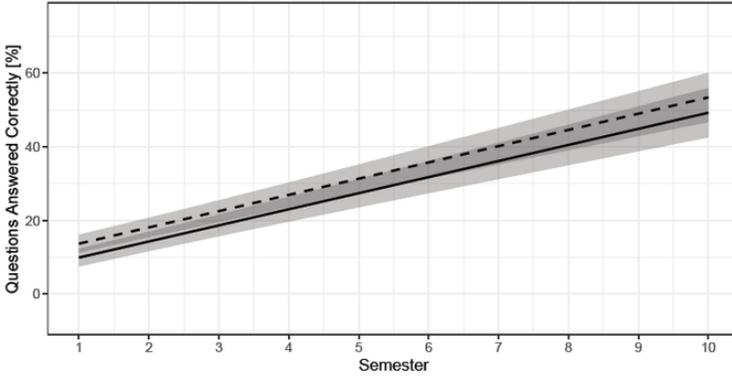
**Figure 1**

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 in figure 1 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, 43 = winter semester 2020/21, respectively). For each semester and PT, the number of participants is indicated above the corresponding boxplot. Abbreviation: PT, Progress Test

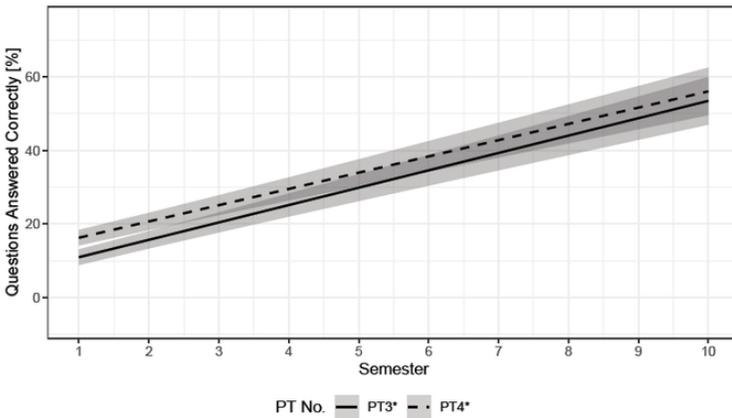
a) Linear Mixed Model for Pair PT36–PT41



b) Linear Mixed Model for Pair PT37–PT42



c) Linear Mixed Model for Pair PT38–PT43

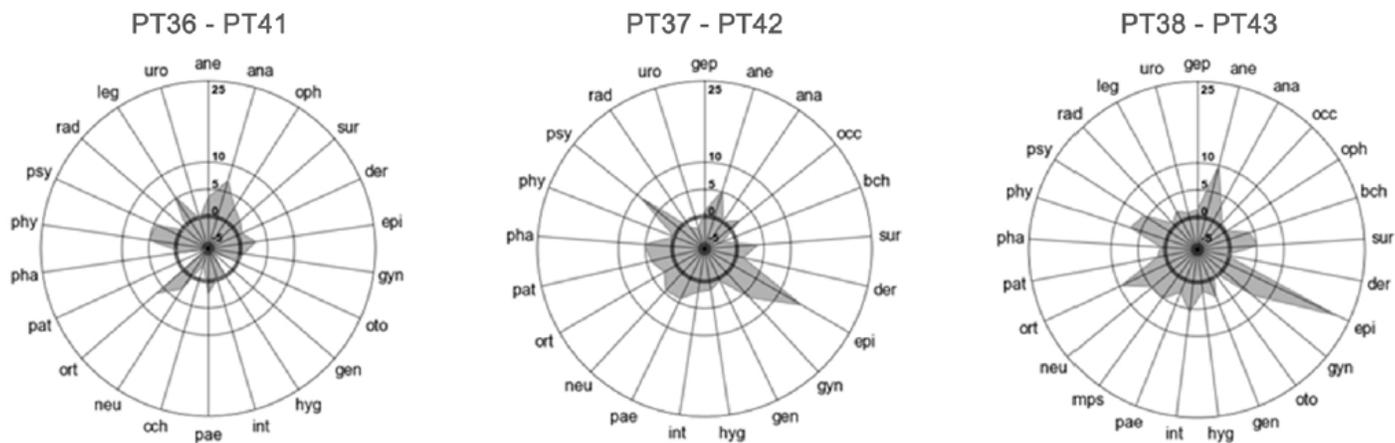


PT No. — PT3\* - - PT4\*

Figure 2

Figure 2 shows the 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. These results refer to the fixed test modality “digital, not ePT”. 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, 43 = winter semester 2020/21); the grey 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. Abbreviations: PT, Progress Test; ePT, electronic Progress Test

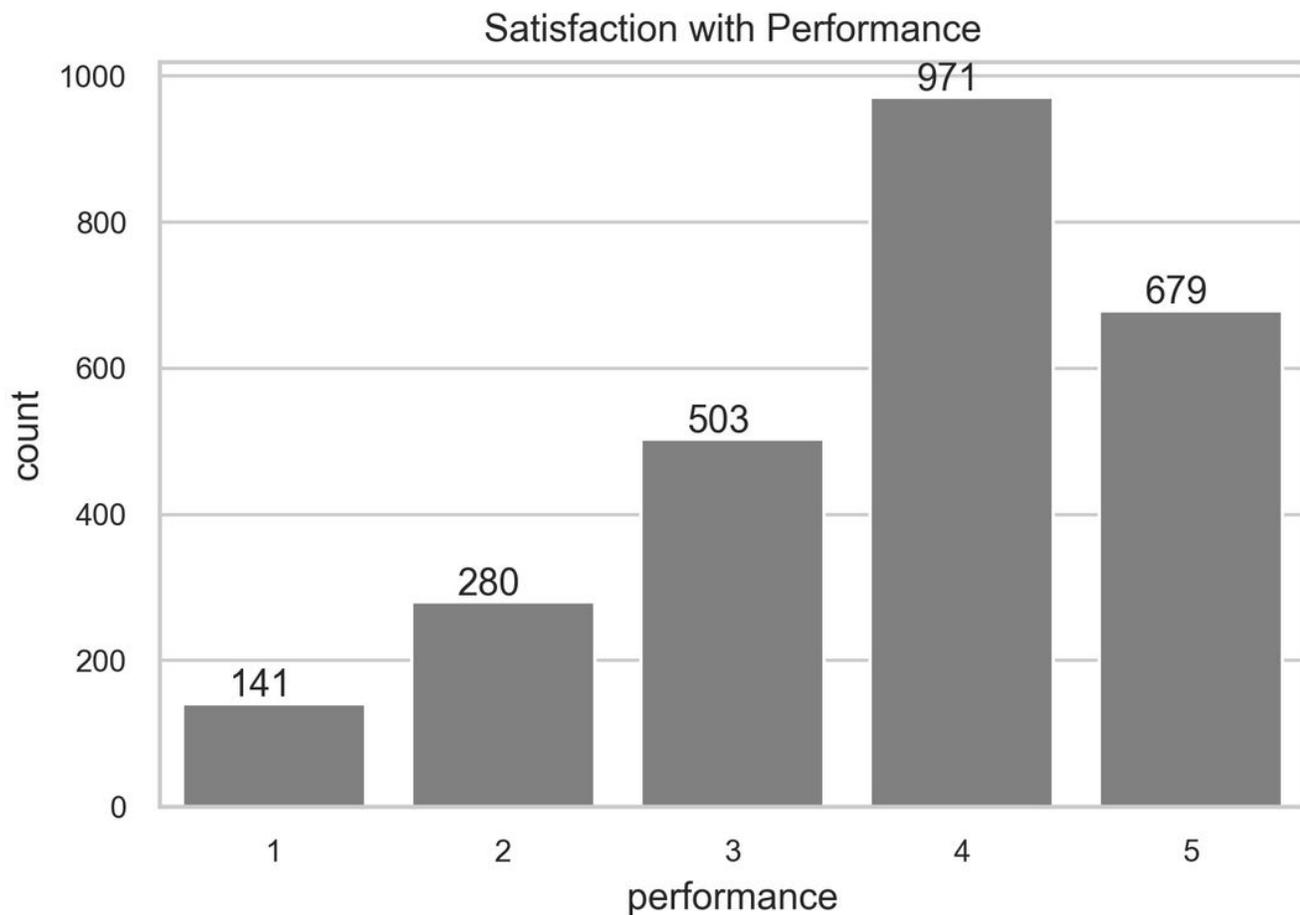
## Student Performance by Subject - Net Changes



**Abbreviations:** gep - General Practice; ane - Anesthesiology, Emergency Medicine and Intensive Care; ana - Anatomy, Biology; occ - Occupational and Social Medicine, Healthcare; oph - Ophthalmology; bch - Biochemistry, Chemistry, Molecular Biology; sur - Surgery; der - Dermatology; epi - Epidemiology, Medical Biometrics; gyn - Gynecology and Obstetrics; oto - Otorhinolaryngology; gen - Human Genetics; hyg - Hygiene, Microbiology; int - Internal Medicine; pae - Paediatrics; cch - Clinical Chemistry, Clinical Pathology; mps - Med. Psychology/Sociology; neu - Neurology; ort - Orthopedics; pat - Pathology; pha - Pharmacology, Toxicology; phy - Physiology, Physics; psy - Psychiatry, Psychosomatic Medicine; rad - Radiology, Nuclear Medicine; leg - Legal Medicine; uro - Urology.

**Figure 3**

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). Abbreviation: PT, Progress Test



**Figure 4**

This bar chart shows that students ( $N = 2,574$ ) from eleven different faculties in Germany and Austria were quite satisfied with their performance in the summative examinations which took place during the academic semesters affected by the pandemic (summer semester 2020 and winter semester 2020/21). They were asked to choose among 5 smileys ranging from very sad = 1 to very happy = 5 (Mean = 3.7; SD = 1.2). The survey was conducted in January 2021.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementDigitalContent060821.docx](#)