Effects of COVID-19 Pandemic on Progress Test performance in German speaking countries.

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)

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, Medical Students, Online Learning, COVID-19 Pandemic

Posted Date: February 23rd, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-786483/v2>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Original Manuscript

**Effects of COVID-19 Pandemic on Progress Test performance in
German speaking countries.**

Corresponding author:

Victoria Sehy (M.Sc.), Charité – Universitätsmedizin Berlin, corporate member of
Freie Universität Berlin, Humboldt-Universität zu Berlin, Progress
Test Medizin, Charitéplatz 1, 10117 Berlin, Germany. Correspondence should be
addressed to Victoria Sehy, AG Progress Test Medizin, Charitéplatz 1, 10117
Berlin, Germany; e-mail: victoria.sehy@charite.de

Abstract

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 students in German speaking countries.

Method The development of the knowledge gain of medical students during the pandemic was measured comparing the outcomes of shared questions within Berlin Progress Test (PT) pairs. 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. It is provided to about 11,000 students in Germany and Austria around the beginning of each semester. We analysed 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), 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.

Conclusion: 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.

Keywords: Progress Test, Medical Students, Online Learning, COVID-19 Pandemic

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 measures¹.

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 pandemic². 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}; stagnated⁵; or just changed for specific subjects⁶ 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 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 graduate level, which provides feedback to students on knowledge and knowledge gain during their course of study⁷. 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 to observe the development among 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:

- 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.

2. Materials and Methods

2.1 Setting

The PT is a low-stakes test on medical knowledge assembled every semester by Charité 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), a digital form where students are required to state how confident they are that their answer is indeed true (ePT), and also other digital forms 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 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 was set to 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) (Additional file 3).

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:

- Faculty-specific PT exam regulations must not have changed between the summer semester of 2017 and the winter semester of 2020.
- The faculty must have administered the test every academic semester since summer semester 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 as ‘serious’ or not; participants classified as non-serious are excluded from the calculation of comparison groups since the validity of results would otherwise be jeopardized⁸.

Non-serious participation is presumed when one or more of the following happens:

- The amount of time devoted to completing the test is too short (less than 20 minutes).
- Every single question of the test was either answered with “don’t know” or not answered at all.

- None of the 120 last questions of the test were responded (suggesting that the test was left incomplete with more than half of it yet to be read).
- Answers show a pattern of randomness^{8,9}.

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, 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 and the lmer function from the R-package lme4¹⁰ for fitting the models, additionally, the semi-partial R^2 was calculated for each fixed effect by using the “nsj” method from the r2glmm package¹¹.

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), PT pair 38-43: 5.66 (4.63-6.69) (see figure 2 and online supplement table 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, 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 in the value of R^2 (for the complete model results see online supplement table 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, 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 online supplement table 4.

4. 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.

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; 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.

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.

We must point out that our findings are not in line with mentioned literature on this topic, which highlights negative impacts of the pandemic on academic performance of medical students^{3,4,6}; on the contrary, our results suggest a sustained performance improvement widely spread across comparison groups, participating faculties and medical disciplines.

In this context, one must keep in mind that the PT is a formative test which assesses the students' entire spectrum of medical knowledge. On the other hand, the existing literature mainly focuses on summative examinations of regular study courses taught during the pandemic. Furthermore, we have found that students in their first semester also performed better in the later PT pair; these students wrote their PT in the first weeks of their studies. We are aware that some of the participating medical schools introduced new admission criteria in 2020¹², which may also have contributed to this outcome. 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. These findings are consistent with survey results in literature where medical students have flagged the benefits of online lectures⁶ and with results from Zgheib et al., 2021 who found that the cognitive performance of students remained the same⁵.

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.

Limitations

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 on medium-term or even long-term effects would require a more prolonged monitoring of results.

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.

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. 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.

List of Abbreviations
COVID-19: Coronavirus disease 2019; PT: Progress Test; CI: Confidence Interval; ePT: electronic Progress Test.

Declarations

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, we also 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 outlined the concept and design of the article. IRA & MS developed the study design, prepared and analyzed the data. All authors discussed and evaluated the results. 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.

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>

References

1. Rogowska AM, Pavlova I, Kuśnierz 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. doi:10.3390/jcm9113494
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*. Published online 2020:1-7.
3. Al-Husban N, Alkhayat A, Aljweesri M, et al. Effects of COVID-19 pandemic on medical students in Jordanian universities: A multi-center cross-sectional study. *Ann Med Surg*. 2021;67:102466. doi:10.1016/j.amsu.2021.102466
4. Harky A, Karimaghaei D, Katmeh H, Hewage S. The Impact of COVID-19 on Medical Examinations. *Acta Biomed Atenei Parm*. 2020;91(4). doi:10.23750/abm.v91i4.1048
5. Zgheib NK, Ali A, Sabra R. Experience with forced transition to online learning during the COVID-19 pandemic: Students' cognitive performance and their perceptions of teaching. *Asia Pacific Sch*. 2021;6(3):45-55. doi:10.29060/taps.2021-6-3/oa2377
6. Kim JW, Myung SJ, Yoon HB, Moon SH, Ryu H, Yim J-J. How medical education survives and evolves during COVID-19: Our experience and future direction. *PLoS One*. 2020;15(12):e0243958.

7. Osterberg K, Kölbel S, Brauns K. Der Progress Test Medizin: Erfahrungen an der Charité Berlin. *GMS Z Med Ausbild.* 2006;23(3)::Doc46.
8. 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). doi:10.1007/s10459-019-09928-y
9. 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.* Published online 2018:1-10.
10. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. *J Stat Softw.* 2015;67(1). doi:10.18637/jss.v067.i01
11. Jaeger BC. Package ‘r2glmm.’ *Packag ‘r2glmm.’* Published online 2017:12. doi:10.1002/sim.3429>is
12. KMK. Studienplatzvergabe im Zentralen Vergabeverfahren: Kultusministerkonferenz verabschiedet Entwurf des Staatsvertrags über die Hochschulzulassung. Published 2018. <https://www.kmk.org/aktuelles/artikelansicht/studienplatzvergabe-im-zentralen-vergabeverfahren-kultusministerkonferenz-verabschiedet-entwurf-des.html>

Figure Legends

Figure 1

Flow Chart of data cleaning process and allocation into different PT pairs.

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. 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, 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

Figure 3

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 3 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

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. Abbreviation: PT, Progress Test

Figures

Figure 1

Flow Chart of data cleaning process and allocation into different PT pairs.

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. 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, 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

Figure 3

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 3 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

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. Abbreviation: PT, Progress Test

Supplementary Files

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

- [OnlineSupplement.docx](#)