From enrollment to exams: Perceived stress dynamics among first-year physics students

Simon Zacharias Lahme,[1,*] Jasper Ole Cirkel,[1] Larissa Hahn,[1] Julia Hofmann,[1] Josefine Neuhaus,[1] Susanne Schneider,[1] and Pascal Klein[1]

[1]*Faculty of Physics, Physics Education Research, University of Göttingen, Friedrich-Hund-Platz 1, 37077 Göttingen, Germany*

(Dated: April 9, 2024)

The current dropout rate in physics studies in Germany is about 60%, with the majority of dropouts occurring in the first year. Consequently, the physics study entry phase poses a significant challenge for many students. Students' stress perception can provide more profound insights into the processes and challenges during that period. In a panel study featuring 67 measuring points involving up to 128 participants at each point, we investigated the students' stress perception with the Perceived Stress Questionnaire (PSQ), identified underlying sources of stress, and assessed the self-estimated workload across two different cohorts. This examination occurred mostly every week during the first, and for one cohort also in the second semester, yielding a total of 3,206 PSQ data points and 5,823 stressors. The PSQ data indicate a consistent stress trajectory across all three groups studied that is characterized by significant dynamics between measuring points, spanning from $M = 20.1, SD = 15.9$ to $M = 63.6, SD = 13.4$ on the scale from 0 to 100. The stress level rises in the first lecture weeks, followed by a stable, elevated stress level until the exams and a relaxation phase afterward during the lecture-free time and Christmas vacation. In the first half of the lecture period, students primarily indicate the weekly exercise sheets, the physics lab course, and math courses as stressors; later on, preparation for exams and the exams themselves emerge as the most important stressors. Together with the students' self-estimated workload that correlates with the PSQ score, we can depict a coherent picture of stress perception among first-year physics students. This study enhances the understanding of stress perception and its potential management. Methodologically, it offers an extensive set of valuable reference data. Practically, it identifies discipline-specific stressors providing educators and faculty with critical insights to undertake stress-reducing interventions to tackle high dropout rates.

I. MOTIVATION: HIGH DROPOUT IN PHYSICS STUDIES

In the academic year 2023, around 11,100 students were enrolled in an undergraduate physics degree program at German universities [1]. This figure has returned to the levels observed in the academic year 2011, marking a significant decrease from the intervening years, during which the enrollment numbers ranged between 14,000 and 16,000 new students annually. Of these students, around 39% do not attend any course in person (so-called park students), so only around 6,800 students have truly started their studies in the recent academic year [1]. Given the frequently discussed shortage of skilled workers, particularly in fields related to natural sciences, an issue exacerbated by demographic change, the number of students enrolling is considered notably low. This lack of enrollment in physics is even more dramatic given the high dropout among physics students. The exact rate is difficult to determine and depends on the specific measurement.

A commonly referenced metric, established by the German Centre for Higher Education Research and Science Studies ([2], assesses the proportion of a specific reference cohort of students who successfully complete their degree programs, regardless of whether they graduate in their original subject or from their initial university. According to the latest data, which is based on a comparison of the 2020 cohort of graduates with students who began their studies in 2016/17 or before, 60% of all students enrolled in a physics or geosciences Bachelor study program at German universities did not receive any degree. This is currently the highest rate compared to all other study programs and it has steadily been one of the highest dropout rates in recent years, with an upward trend from 39% among students who started in 2006/07 to 49% in 2014/15 and lately 60% [2, 3]. Another method to estimate the dropout involves comparing the actual number of physics Bachelor graduates (2,660) in 2023[1], which is consistent with figures from previous years [1], with the aforementioned enrollment numbers. This yields a success rate of less than 25%.

For physics teacher training programs, these problems are even more serious as out of around 2,000 students enrolled annually, only around 300 complete the Bachelor's, and slightly fewer the consecutive Master's degree program [1]. Forecasts on the recruitment needs of new teachers show that the number of new graduates is far below the actual demand [6].

Low numbers of natural sciences students and high dropout are also international issues. Natural sciences, mathematics, and statistics have the lowest share of graduates in OECD countries compared to all other fields [7], so physics is a very narrow field of study [8]. Data from the late 1990s and early 2000s show that high dropout in physics has already been an issue for a long time in various European countries [9].

In about half of the cases, the dropout occurs already in the first year of study [10], highlighting the necessity of focusing research efforts on understanding the underlying processes in this physics study entry phase. Here, stress perception can be considered as a pivotal element for better understanding students' experiences that could lead to them dropping out (e.g., difficulties, challenges, circumstances). Therefore, this paper examines students' stress perception in that particular phase.

* simon.lahme@uni-goettingen.de

[1] For comparison: In the US with around four times more inhabitants than Germany, 9,031 students graduated from a physics Bachelor program in 2020 [4]; in the UK with around 20% less inhabitants than Germany, 3,675 students graduated from an undergraduate physics course in 2015 [5].

II. STATE OF RESEARCH AND RESEARCH QUESTIONS

A. Research on the physics study entry phase

Study dropout rates and underlying causes have already been investigated and discussed in Germany for more than 40 years [11]. Besides descriptive trend studies for all subjects like the mentioned studies from the DZHW, there has also been subject-specific research for science and physics study programs in particular. In this context, three strains of research can be identified: research on risk (1) and success (2) factors as well as efforts to innovate the study entry phase (3).

The first strain on risk factors investigates the extent of dropout, underlying processes, and reasons. Ref. [12], e.g., has examined risk factors for a successful start in a physics study program. The study is based on work by Ref. [13] who developed a model for academic success based on the students' individual study and learning behavior that depends on personal entry preconditions (e.g., prior knowledge from school, study choice motives, or socio-demographic background), study conditions (e.g., quality of university teaching, the structure of the study program, or organization), and more general context conditions (e.g., employment, illness, or family situation). Ref. [12] considered aspects like the grade of the university entrance qualification, the supervision, support, insufficient information, and interest in the subject and identified content requirements as the most salient reason for exmatriculation from a physics study program.

The second strain investigates factual and affective aspects of the physics study entry phase and their impact on academic success. Examples are the students' prior knowledge in mathematics and physics [14–16] and their overall study ability [17], their acquisition of factual knowledge and problem-solving competencies during the introductory phase [18, 19], the students' views of the nature of science [20], or their sense of belonging to the physics community and university [21]. Similar research has also been conducted in other countries, focusing on topics like the importance of math prerequisites [22] and the sense of belonging in the US [23, 24], or emotions of physics students in Australia [25] and Finland [26].

The third strain, innovation, is about the improvement of the physics study entry phase to better support the teaching and learning process and facilitate overall academic success. Typical innovations are the implementation of extracurricular measures to tackle or compensate for discrepancies between the actual and expected students' performances. For example, learning centers have been established to provide supportive learning offers in addition to the regular curriculum (e.g., [27]). Pre-courses have been introduced as on-campus courses and online learning units so that students could repeat relevant school mathematics bridging the gap between school and university mathematics [28–32]. Additionally, efforts have been dedicated to the intracurricular enhancement of physics study programs. Examples include the implementation of classroom exercises, complexity-graded tasks, and supportive learning materials for subject-specific problem-solving [33–35], or smartphone-based experimental tasks [36–39]. A review of innovations' impact on learning in undergraduate science courses is available in Ref. [40].

B. Perceived stress & stressors in university education

1. *Definition of stress*

Building on the idea of the first strain, the investigation of risk factors for study dropout, the exploration can extend to students' stress perception and the underlying causes of stress. Psychological stress can be defined as "a particular relationship between the person and the environment that is appraised by the person as taxing or exceeding his or her resources and endangering his or her well-being" [41, p.19]. Stress perception is based on a cognitive appraisal process "of categorizing an encounter, and its various facets, with respect to its significance for well-being" which is of "largely evaluative" and continuous nature [p.31]. The appraisal process takes into account both person and situation factors and particularly depends on the availability of strategies for coping, i.e. the "constantly changing cognitive and behavioral efforts to manage specific external and/or internal demands that are appraised as taxing or exceeding the resources of the person" [p.141]. As stress perception is a very individual interplay between a person and their environment, the same stressor, i.e., "a subset of environmental conditions that are likely to be appraised as demanding and to have implications for a person's well-being" [42, p.30] can be perceived differently by different people.

2. *Measurement of stress perception with the PSQ*

A rather popular instrument for measuring stress perception is the *Perceived Stress Questionnaire* (PSQ), originally conceived for the investigation of psychosomatic patients [43]. It has been translated into various languages like German [44], Spanish [45], or Chinese [46, 47]. The different language versions differ in the number of items and subscales. The German version comprises 20 items (statements), and a 4-point Likert-type rating scale to assess the leve of agreement with each statement, viz. *almost never* (1), *sometimes* (2), *often* (3), and *most of the time* (4). They form four subscales with five items each, which describe the perceived *worries*, *tension*, and *joy* that seem to reflect the internal stress reaction of the individual, and *demands* that seems to refer to the perception/appraisal of external stressors [44]. Worries, tension, and demands are treated as increasing while joy is treated as decreasing the total perceived stress score that is derived as the mean across all items after partial inversion of items and is scaled linearly from 0 (min.) to 100 (max.).

A broader body of research exists where the PSQ has been employed to investigate students across various disciplines (cf. Table I). It was used in the fields of medicine, nursing, and health, particularly to validate the different versions of the PSQ, to investigate education students, or to get an overview of stress perception at universities, e.g., during the COVID-19 pandemic by investigating students in various subjects and years of study. Data was collected on different time scales from a few days to several months and in always one measuring period, except from Ref. [48] who measured three times at the beginning, mid, and end of a semester. The sample sizes, means, and standard deviations vary considerably.

TABLE I. Overview of literature providing comparative data from the Perceived Stress Questionnaire (PSQ) for students. For each reference, it is indicated which version of the instrument was used, i.e. the number of items (13, 20 or 30), the language (Chinese (C), English (E), German (G), Spain (S)), and whether a four- or six-point rating scale was used. As far as identifiable, a description of the investigated students and measuring points is provided together with the number of students (N) and their mean total perceived stress score (M) including standard deviation (SD) and error (SE). For better comparability, all perceived stress scores are presented on a scale from 0 to 100 for what data partly needed to be transformed. Sometimes, e.g., the total stress score needed to be determined based on subscale values. In that case, the average of the subscale values as well as the highest standard deviation were used for approximation.

Reference	Instrument	Investigated students	Measuring point	N	M	SD	SE
[48]	PSQ-20-G-6	Across all subjects & years of study including PhD, University of Nürnberg-Erlangen, Germany	begin of summer semester 20	2795	42.10	22.80	0.43
			mid of summer semester 20	2795	48.60	23.00	0.44
			end of summer semester 20	2795	50.50	23.40	0.44
[49]	PSQ-20-G-6	Education (primary & secondary school), University of Nürnberg-Erlangen, Germany, various years of study	begin of summer semester 20	51	39.00	27.80	3.89
			mid of summer semester 20	51	49.90	25.40	3.56
			end of summer semester 20	51	51.30	27.00	3.78
[44]	PSQ-20-G-4	Medicine, Germany, 4th or 5th clinical semester		246	34	16	1.02
[50]	PSQ-20-G-4	Across all subjects & years of study, several universities in Germany	12/11-01/12	2435	52	19	0.39
[51]	PSQ-20-G-4	Medicine, University Medical Center Hamburg-Eppendorf, Germany, 1st year of study	01/14 during regular seminar, at the beginning of the second six-week module before the written semester exam in February	321	40	15	0.84
[52]	PSQ-20-G-4	Education, part of an introductory psychology lecture, University of Bamberg, Germany, various years of study	one measurement in winter semester 18/19, summer semester 19, winter semester 19/20, & summer semester 20	110	40.75	23.6	2.26
[53]	PSQ-20-G-4	Medicine, part of a communication course, Germany	winter semester 20/21	136	49.53	10.64	0.91
[54]	PSQ-30-E-4	Health & activity course, Mid-Atlantic University, US, undergraduate level		232	62.8	12.7	0.83
[55]	PSQ-30-E-4	Across all subjects & years of study, University of York & Hull York Medical School, UK	05/20-06/20	788	51	20	0.71
[56, 57]	PSQ-30-S-4	Dental, Universities of Uesca & Santiago de Compostela, Spain, various years of study	05/11, two weeks before final exam phase	314	45	19	1.07
[58]	PSQ-24-S-4	Education, Catholic University of Valencia, Spain, undergraduate level	end of classes, 03/16-05/16	589	42	17	0.70
[46, 59]	PSQ-30-C-4	Medicine, University of Wuhan, China, post- & undergraduate	12/16-01/17	122	40.2	13.3	1.20
		Medicine, University of Ningbo, China, junior college	11/15-01/16	1453	39.9	13.8	0.36
[47]	PSQ-13-C-4	Medicine, University of Hangzhou, China, various years of study	09/21-12/21	309	29.019	6.325	1.944

3. *Further research on students' stress perception*

Beyond the use of the PSQ, there have been further studies about students' stress perception. Some of them do not explicitly focus on a specific subject [60–65], others address specific subjects like nutrition [66], teacher training students [67–69], or chemistry [70]. Research interests, objectives and methods vary widely in these studies and this list is not exhaustive. Therefore, we only highlight three studies that are particularly relevant in our research context.

Ref. [60] used the Perceived Stress Scale (PSS) by Ref. [71] among German students in two weeks in late November 2011 leading to the observations that 59% of the students felt nervous or stressed within the past weeks and 31% had the feeling that they could not control important aspects in their life. No relevant differences between subjects were found but women and students in higher semesters tend to perceive more stress. Stress was mostly associated with time and performance pressure (in tendency more in earlier years of study) as well as fear of the future and uncertainty (in tendency more in higher years of study), and excessive demands posed on students. In the study, students also rated eleven different areas of life (studies, work, family situation, financial situation, leisure, health, household, children, partnership, social contacts, and housing situation) identified by Ref. [72] according to which extent they contribute to their stress perception. The findings show that on average students perceive three areas of life as strong stressors. The most frequent stressors are the studies themselves (68% of the students) as well as the financial situation, work, and leisure (around 40% each).

Similarly, Ref. [65] investigated the relevance of ten predefined stressors as perceived by 243 students of various subjects and years of study at a German university. The responses reveal that general conditions of the university like rooms, facilities, or number of participants in courses, requirements regarding the course of studies and timetable in the major subject, the individual financial situation, and the working style as a fit between habits and external requirements are the most important stressors. Requirements in the minor subject, conflicts between the studies and private interests, and jobbing are perceived as medium important. The need for active participation in the courses, troubles with the self-concept of ability, and family obligations are rather unimportant stressors for most students.

Ref. [70] investigated the stress of 178 first-semester chemistry students at two German universities following the same definition of stress by Ref. [41]. Students monthly rated on a five-point Likert scale a single item by Ref. [73] to what extent they felt that they could meet the demands of their study program in the same week. Furthermore, they could state reasons for their feeling of overload in an open text field. The quantitative data show that 38% of all students perceive a moderate and 26% a high or very high mismatch (similar to the level of stress) during their first year of studying chemistry. The analysis of the qualitative data leads to four areas of stressors. The most prominent area is built by cognitive-factual stressors which were split into qualitative (i.e., level of difficulty), quantitative (i.e., workload), and not further specifiable stressors as well as responses regarding exam failure. Furthermore, cognitive-organizational, physical (including sickness), and social stressors were identified but were less common. While qualitative stressors dominated the first half of the first semester, the second half of the semester was dominated by quantitative stressors due to the lab course and exam preparation. The math and lab course and particularly the weekly exercises and self-study at home were the most relevant aspects linked to these quantitative and qualitative stressors while exercises and tutorials were hardly mentioned.

4. *Stress and the role of academic workload*

An accompanying factor that is sometimes investigated together with students' stress perception is the academic workload. For instance, Ref. [50] utilized the PSQ alongside an item by Ref. [74], where students were asked to report the number of hours they spend in an average lecture week on attending courses, engaging in course-related activities, and working to finance their studies. The analysis shows that the self-estimated workload explains a significant part of the variance in the PSQ sub-dimensions tension and demands showing that workload is an interesting variable to consider together with the PSQ. The average workload reported in that study is comparable to a typical full-time job as around half of the students stated a workload between 31 h and 50 h per week, with extreme values even beyond 70 h.

Conversely, an extensive time budget survey [75] across various subjects and German universities required students to detail their daily timetables, providing in-depth information on activities regarding their studies as well as estimates on private time, jobbing, vacation, illness, etc. Findings show that students spent significantly less time on their studies than one would expect. Based on the EU-wide ETCS system, students should spend around 1,800 h per year on their studies, i.e., 45 weeks á 40 h (and 7 weeks holidays) equal to on average 34.6 h per week or 4.9 h per day. However, at no university and in no study program investigated by Ref [75], this 4.9 h/d limit is exceeded in any month of the semester (besides one minor exception). Usually, this limit is even significantly undercut (with the lowest value 1.4 h/d) leading, on average, to a 23 h week significantly below the 34.6 h. The authors point out the comparability of their own findings with those from two American studies [76, 77] which have shown that the workload of four-year college students decreased from 40 h/week in 1961 to around 27 h/week in 2003 [78].

C. Research questions

The described state of research indicates a considerable interest in examining students' stress perception across various subjects and countries. Yet, it also highlights a significant gap: Despite physics study programs having the highest dropout rate among all study programs in Germany, there appears to be no study specifically focused on the stress perception of physics students as it has already been done for many other subjects. Furthermore, many studies have only one or very few measuring point(s) that do not allow a high resolution of possibly time-dependent stress perception and stressors in the first year of study. Moreover, the PSQ seems to be an established and frequently validated instrument to investigate stress perception of university students. In response to the identified need to understand physics students' perception of stress and stressors in the study entry phase, we pursue the following research questions:

RQ1: *How does the perceived stress of physics students (measured by the PSQ) evolve during the first year of their study?* The first question deals with the temporal dependency of the stress perception of physics students in their first year of study. An answer to this question will provide quantitative data on which phases of the study entry phase are perceived as more or less stressful by the students, as well as an overall assessment of their stress level.

RQ2: *How does the self-estimated workload of the students evolve during the first year of their study?* The second question is a supplement to the first and deals with the students' self-assessed weekly study-related workload as another indicator that could be related to stress perception.

RQ3: *What are the most salient stressors contributing to physics students' perceived stress, and how does their indication shift during the first year of their studies?* The third question goes beyond the quantitative data by looking at the stressors that contribute to perceived stress from the students' point of view, and as above, the temporal dependency of these stressors over the semester is investigated. An answer to this question provides a deep insight into the (subject-specific) causes of stress perception. This will be an important basis for future supportive measures and systemic improvements.

In addition, we investigate whether there are differences between different cohorts and between first and second semester students. This will show whether the stress perception depends on the cohort and associated circumstances (e.g., at the university) and whether the first semester differs from the second semester, in which students are already more familiar with the university and their study program.

III. METHODS

A. Data collection and sample

The stress perception was studied among first-year, i.e. first and second semester physics students in a voluntary panel study at the University of Göttingen. As Figure 1 shows, the students attend, in both their first and second semesters, the courses *Experimental physics* and *Mathematics for physicists* (or alternatively *Integral and differential calculus*). Additionally, they attend *Mathematical Methods in Physics* in the first and *Analytical Mechanics* in the second semester. Each course consists of one or two lectures, a tutorial, and an exercise sheet every week and ends with a written exam. A lab course accompanies the experimental physics courses. In the lecture-free time between the first and second semester, there is a programming course. Before the first semester, students can optionally attend a pre-course and an orientation week for onboarding. The structure of the physics teacher program is slightly different as they have a second study subject.

The survey was conducted in three consecutive semesters as part of the basic courses *Experimental Physics* and *Mathematical Methods in Physics*. In winter semester 21/22 (10/21-03/22) and summer semester 22 (04/22-09/22), data was collected from a cohort A of students in their first and subsequent second semester, in the following referred to as groups A1 and A2. In winter semester 22/23 (10/22-03/23), the next cohort B of first-semester students was studied, in the following referred to as group B1. Examining these three groups allows us to compare how stress perception differs between the first and second semesters and between two different cohorts.

As far as possible, data was collected via an online survey tool during the lectures within a short break, so that the students could participate in the survey immediately and repeatedly. This mode of data collection was possible since one of the co-authors (PK) was the responsible lecturer in all investigated courses. Depending on the response rate during the lectures and in particular beyond the lecture times, the online survey link was sent to the students via e-mail to ask for participation. For group A1 during lecture times, additionally, a paper version was used in the beginning. However, for economic reasons and to enable data collection from students not regularly present on campus (e.g., during lecture-free periods or due to the COVID-19 pandemic), we subsequently transitioned to online surveys. This approach was adopted for all measurement with groups A2 and B1.

As shown in Figure 2, in the winter semesters (groups A1 and B1), data was collected weekly in the two pre-course weeks and the orientation week, in each of the 14 lecture weeks, during the Christmas break, and partly bi-weekly in the exam weeks and the lecture-free weeks thereafter. In summer semester 22 (group A2), data was collected during the 14 lecture weeks, once during the exam period, and partly bi-weekly in the lecture-free weeks thereafter; there was no pre-course or break during lecture time in that semester. This data collection scheme resulted in a total of 67 measuring points, 23 for group A1, 21 for group A2, and 23 for group B1.

Pseudonymous codes were used to link responses from different measuring points. Potentially misspelled codes were matched if three out of four code elements were the same and if there was no obvious reason why these codes could not belong to the same person. Table II shows the number of matched codes, i.e., the number of participating students in the three groups and available demographic data. 116 to 164 codes were identified per semester, and demographic data is given for 88 to 134 codes. On average, 87% of the participants were in their first respectively second semester, 75% physics majors, 18% physics teacher training, and 7% other students. 24% were female. Overall, most of the registered participants can be attributed to our primary target group, physics Bachelor and teacher training students in their first or second semester, so we subject the entire dataset to further analysis. The number of participants for each measuring point is given in Figure 4. The complete quantitative dataset is available as supplemental material.

B. Instrument

1. Overview, structure, and items

The main instrument (cf. Table V in the appendix) is the German version of the PSQ [44] in its original formulation. Although it has not been used specifically for first-year physics students yet, the questionnaire has already been validated and used for students of other subjects. All items can also be interpreted within the context of studying physics, e.g., "*You feel that too many demands are being made on you.*" (worries) or "*You feel under pressure from deadlines.*" (tension). To provide a reference point, the questionnaire instructed the students to relate the items to their current situation of studying physics. Like Refs. [48, 79], we used a six-point rating scale instead of the original four-point rating scale to enhance resolution for repeated measurements. In the online survey, items were presented in randomized order. In addition to this slightly adapted PSQ, students should assess the workload they had spent on their studies in the past week and were asked to indicate up to three stressors that are currently most relevant to their stress perception in open-text fields.

2. Instrument characterization

In Figure 3, the distribution of the total perceived stress score across all participants and measuring points is displayed ($N = 3,206$ observations). It shows that students reported perceived stress across the entire scale from 0 to 100, with an average of $M = 50.3$ ($SE = 0.34,\ SD = 19.5$). Visual inspection of empirical and theoretical cumulative distribution function (cf. Figure 11 in the appendix) and D'Agostino test of skewness ($skew = -.078,\ z = -1.80,\ p = .071$) show that this distribution can be considered normal.

To assess the internal consistency, Cronbach's alpha was calculated for a focus sample, Pre1 of groups A1 and B1 ($N = 196$ students), because these two measuring points are considered the least influenced by the study program as data was collected only a few days after students entered university and the actual lecture time had not started yet. It is $\alpha = 0.89$

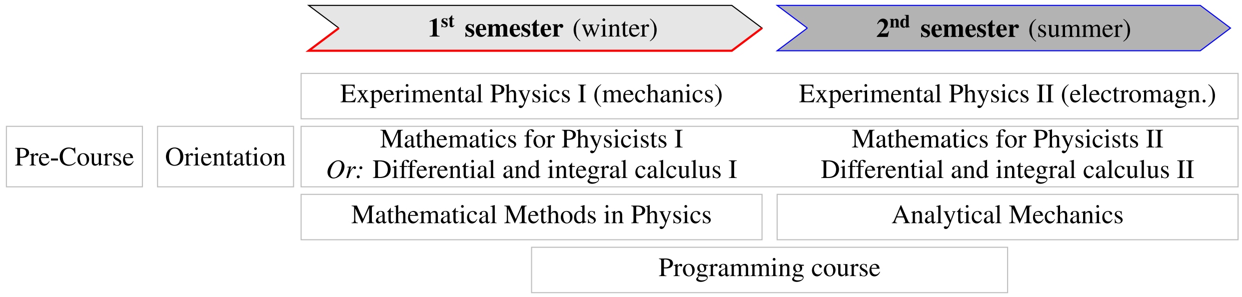


FIG. 1. Structure of the first and second semesters of the physics Bachelor study program at the University of Göttingen.

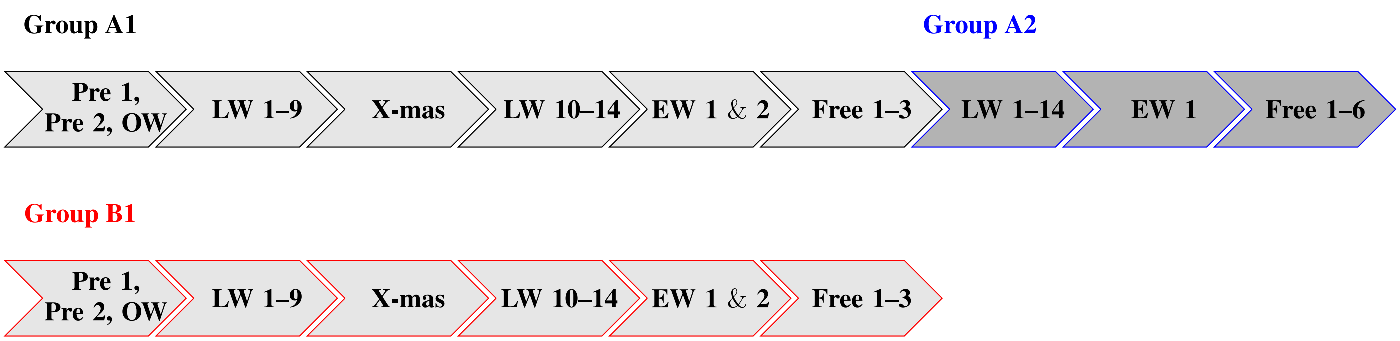


FIG. 2. Overview of the data collection procedure for groups A1, A2, and B1 in the 1st semester (light gray) and 2nd semester (dark gray). (Pre = precourse, OW = orientation week, LW = lecture week, EW = exam week, Free = lecture-free week, and X-mas = Christmas break)

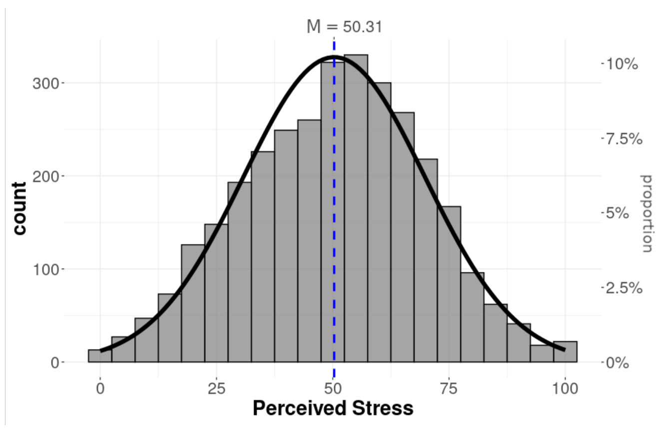


FIG. 3. Total score of perceived stress distribution based on the complete dataset ($N = 3,206$) with a normal distribution ($M = 50.3,\ SD = 19.5$); this includes multiple responses from the same student at different measuring times.

which is sufficiently satisfactory and in accordance with reference values by Ref. [44] with $\alpha = 0.83$ for $N = 246$ medicine students at the beginning of their semester and $\alpha = .85$ for the total sample group ($N = 650$) in that study.

3. *Data interpretation*

A common approach for interpreting PSQ data, as employed by Refs. [51, 80, 81], is based on the mean and standard deviation of the usually single measuring point. A total score $\leq M + 1\ SD$ is interpreted as a *mean* stress level, a score $> M + 1\ SD$ as a *slightly increased* stress level, and $> M + 2\ SD$ as a *high* stress level. To allow for a similar interpretation of the Göttingen data while accounting for repeated measurements, we derive these intervals from the focus sample with a total stress score of $M = 35.7,\ SD = 14.7$. Consequently, we interpret total stress scores between $[0, 50]$ as a *mean*, between $(50, 65]$ as a *slightly increased*, and between $(65, 100]$ as a *high* stress level.

C. Method of data analysis

1. *Quantitative data - stress perception & workload (RQs 1 & 2)*

Since not every student responded to the survey at every measuring point, PSQ and workload data are further processed for statistical analysis. For this, we focus on the precourse, orientation week, and lecture weeks due to the significantly lower participation rates during Christmas breaks, exam weeks, and lecture-free weeks (cf. Figure 4). This is also supported content-wise as the lecture time is the most relevant phase during the semester and its requirements and circumstances are more similar among the students compared to the lecture-free time. To further deal with missing responses during the selected measuring points, we applied data smoothing by combining three consecutive weeks, resulting in five or six corresponding measuring periods for the summer and winter semesters, respectively: Period 0 is Pre1, Pre2, and OW,

TABLE II. Sample overview. The total number of participants is the aggregation of all different codes registered at any measuring point. Some codes were provided with slight variations; these were matched and considered as a single code to ensure accuracy in data representation.

	Group A1	Group A2	Group B1
Investigated group	Cohort A, 1st semester	Cohort A, 2nd semester	Cohort B, 1st semester
Investigated course	Mathematical Methods in Physics in winter semester 21/22	Experimental Physics II in summer semester 22	Experimental Physics I in winter semester 22/23
Total number of registered, matched codes	145	116	164
Demographics available	107	88	135
of which are in 1st resp. 2nd semester	97	64	125
of which are physics major	83	66	99
of which are teacher training students	18	15	25
of which are female	25	26	28
of which are of divers/unstated gender	3	1	7

period 1 is LW1-LW3, period 2 LW4-LW6, period 3 LW7-LW9, period 4 LW10-12, and period 5 LW13-LW14. They are gray-shaded in Figures 4 and 5. For each period, we average up to three available total stress scores gathered during these periods individually for each participant. Additionally, if students did not participate in only one single measuring period, the mean of the corresponding measuring period is imputed for the missing value. In total, we obtain 52 complete datasets, i.e., five/six stress scores representing all measuring periods in the pre-lecture and lecture time for group A1, 54 for group A2, and 76 for group B1. When matching groups A1 and A2 being the same cohort of students, 39 complete datasets remain. The same procedure for smoothing the data was analogously applied to the workload data.

The processed datasets (cf. Figures 12 and 13 in the appendix) were statistically analyzed to compare stress perception and workload between the three groups and to examine the progression over the semester. A mixed ANOVA was used to compare groups A1 and B1 (cohort A vs cohort B) with the group as between-subject and six measurement periods as within-subject factor. Repeated measure ANOVAs were used to compare groups A1 and A2 (first vs second semester) with eleven measurement periods for the matched sample of $N = 39$ students and to analyze the changes over the five/six measurement periods in the first and second semesters. According to Shapiro-Wilk tests for each group of interest, data is normally distributed in most periods but extreme outliers were found in several periods for the workload data, so a 90% winsorization was applied for workload data. For the mixed ANOVA, the homogeneity of error variances and covariances was checked by Levene's and Box's test which was given except the homogeneity of covariances for the total stress score ($p = 0.011$). For the ANOVAs, Greenhouse-Geisser adjustment was applied due to the lack of sphericity shown in Mauchly tests. In case of significant temporal evolution, a Bonferroni-adjusted post-hoc analysis was conducted.

2. *Qualitative data - stressors (RQ3)*

Across all measuring points, students indicated 5,823 stressors, 2,216 by group A1, 1,248 by group A2, and 2,359 by group B1. The responses were subjected to a structuring qualitative content analysis leading to a category system summarized in Table III; the full category system is available as supplemental material. The analysis of the codings reveals the most salient stressors in different phases of the semester.

The category system consists of three dimensions, each subdivided by categories and, in some cases also subcategories. These dimensions describe whether the stressors are primarily related to the *university* and the students' studies (U), primarily *private* (P), or *globally* affect both the university and private sphere (G). The *university*-related dimension is further divided into eleven categories. Some of them describe typical activities of studying physics, i.e., writing *lab reports* (U6), *preparing and following up the lectures* (U7), understanding the *lecture contents* (U8), solving regular *exercise sheets* (U9), or *preparing for and writing exams* (U10). Further categories acknowledge the *conditions of study* (U1) for which the university holds responsibility, challenges related to the *transition from school to university in the first semester respectively the transition from one semester to the other* (U2), the overall *organization of one's own study program* (U3), the *individual everyday study routine* closely linked to the *time management* (U4), and the *study-related self-regulation* (U5) about motivation, emotions, thoughts, doubts, etc. Moreover, there is a category designated for all *courses and subjects that are mentioned without specifying* (U11) the precise underlying stressors involved. There is also a category designated for a *project group work* (U12) that was newly introduced in the winter semester 22/23. The *private* dimension is divided into three categories addressing *everyday demands* (P1) like household, private activities, etc., the *private social environment* (P2) and relationships as well as *illness* (P3). The *global* dimension is divided into four categories related to the *financing of one's studies* (G1), fears/uncertainties about general *future prospects* (G2), the *COVID-19 pandemic* (G3), and the *work-life balance* (G4) including relaxation. In an additional miscellaneous dimension/category, all further responses can be categorized that cannot be definitely assigned to the previous categories due to their lack of specificity or rarity.

To assess the reliability of the category system, an interrating was conducted involving the first author (rater 1) and three 2nd to 3rd-year teacher training students (raters 2 to 4) to integrate the student's perspective in the data analysis. Around 31% of the responses were subjected to the interrating, with careful attention to ensure a balanced distribution of data across different measuring points in the semester and the three groups. Initially, each rater independently coded sub-

TABLE III. Overview of the category system for the stressors listed by the students in open text fields. It consists of three dimensions (university, global, and private) and categories plus miscellaneous. The full category system including subcategories, more detailed descriptions, and anchor examples is available as supplemental material.

Category		Short description
University		
U1	Conditions of study	General (study) conditions at university, in particular, university guidelines, structures, (information) offerings as well as the university's personnel and material resources
U2	Transition from school to university or start of semester	Transition from school to university or the transition from the semester break to the new semester, for example, a lack of prior knowledge, catching up on school material or material from the previous semester, or finding study groups
U3	Study organization	Individual organization of one's own studies, i.e., planning and structuring a semester or the entire course of study
U4	Individual everyday study routine/time management	Daily study routine, in particular individual weekly planning and time management
U5	Study-related self-regulation	Fears, insecurities, or pressures directly related to one's studies, including motivational factors and strategies
U6	Lab reports	Writing the lab reports/protocols as part of the physics lab course: requirements, number, scope, or deadlines, or unspecific
U7	Preparation and follow-up of lectures	Any work in addition to exercise sheets, etc. that serves to improve understanding of the lecture content and close gaps in knowledge
U8	Lecture contents	Subject matter in the lectures: novelty, difficulty, quantity, pacing, or unspecific
U9	Exercise sheets	Weekly exercise sheets (problem sets): difficulty, quantity, or unspecific
U10	Exams & exam preparation	Exam admission & exam preparation, exam phase in particular, or fear/anxiety of exams
U11	Unspecific mention of courses & subjects	Unspecified stress caused by their subject or courses attended, divided after math, physics, and lab courses, the pre-course, and other subjects/courses
U12	Project group work	Stressors related to undergraduate research projects newly introduced only for group B1
Global		
G1	Financing of studies	Burden concerning the own financing, including employment and general financial worries
G2	Future prospects	Study-unrelated (unspecific) fears, worries, and uncertainties regarding the prospects for the present and future
G3	Covid-19 pandemic	Burden that is directly related to the Covid-19 pandemic
G4	Work-life balance	Strain on one's personal work-life balance, referring to private opportunities for relaxation or, conversely, when private obligations interfere with studies (in terms of time)
Private		
P1	Everyday demands	Burden caused by the demands of daily everyday life, including factors and especially activities related to (new) household management and living arrangements as well as extracurricular commitments, activities, and hobbies
P2	Private social environment	Social contacts in their personal environment or the establishment of a new social environment, including family, friends, and private contacts
P3	Illness	Temporary or long-term, physical or psychological illness including unspecific descriptions of a poor state of health
Miscellaneous		Stressor that does not fall under one of the other categories or cannot be meaningfully assigned

samples of responses. Then, differences in the ratings were discussed in the group of each three participating raters. Table IV presents Cohen's kappa for two and Fleiss' kappa for three raters before (κ_{pre}) and after (κ_{post}) discussion of codings on the subcategory level of the category system. All κ_{pre}-values, except the Fleiss' kappa for group B1, exceed the threshold of 0.61 indicating *substantial* agreement; all κ_{post}-values surpass the threshold of 0.80 and can therefore be interpreted as *almost perfect* [82]. The discussions led to various minor refinements of the category system including clarifications and delimitations between the categories that have been implemented in the current category system and codings.

TABLE IV. Cohen's κ for two raters and Fleiss' κ for three raters, measure the agreement on approximately 31% of the mentioned stressors coded using the category system before (κ_{pre}) and after (κ_{post}) discussions among the raters of the interrating. Rater 1 was the first author, and raters 2 to 4 Bachelor teacher training students. According to Ref. [82], values between .61 and .80 are *substantial*, higher values are *almost perfect*.

Comparison	Group A1		Group A2		Group B1	
	κ_{pre}	κ_{post}	κ_{pre}	κ_{post}	κ_{pre}	κ_{post}
Rater 1 vs 2	0.80	0.92				
Rater 1 vs 3	0.78	0.92	0.76	0.97		
Rater 2 vs 3	0.76	0.91			0.71	0.93
Rater 1 vs 2 vs 3	0.71	0.88				
Rater 1 vs 4			0.79	0.95		
Rater 3 vs 4			0.74	0.96	0.70	0.95
Rater 1 vs 3 vs 4			0.68	0.95		
Rater 2 vs 4					0.72	0.94
Rater 2 vs 3 vs 4					0.59	0.90

IV. RESULTS

A. Trajectory of perceived stress (RQ1)

Figure 4 shows the trajectory of perceived stress, i.e., the average stress level of all participating students for each measuring point on the scale from 0 (min.) to 100 (max.) over the semester for the groups A1, A2, and B1. Apparently, the three trajectories exhibit a similar pattern. In the winter semesters (groups A1 and B1), students start with a perceived stress level of around 35 in the pre-course. In the first four to five lecture weeks, the stress level increases in all groups to a level of around 55, a level sustained with minor fluctuations until the exam weeks. After the exams, the stress level decreases even below the starting level down to around 20 to 30. In group A1, the stress perception also diminishes during the Christmas break. In group A2, measuring points in the later lecture-free time indicate a minor increase in the stress level.

1. Differences between the three groups

For groups A1 and B1, both cohorts in their first semester, a mixed ANOVA supports the impression that the course of stress perception is similar for both groups. There are neither interaction effects between the measuring period and group ($F(4.2, 526.1) = 1.3$, $p = 0.26$, $\eta_P^2 = 0.01$) nor differences between the groups ($F(1, 126) = 0.5$, $p = 0.48$, $\eta_P^2 = 0.003$). Thus, we merge the datasets of groups A1 and B1 when analyzing the temporal progression of perceived stress throughout the first semester in the next section.

For the matched sample of groups A1 and A2, a repeated measure ANOVA shows a significant temporal evolution of the perceived stress within the two semesters ($F(5.7, 215.6) = 40.0$, $p < 0.001$, $\eta_P^2 = 0.51$). Post-hoc analysis, where we first focus on a comparison of corresponding measurement periods, shows different levels of perceived stress between the semesters during period 1 but equal stress levels thereafter. The total stress score in period 1 of group A2 is significantly lower than the score in period 1 of group A1 ($p = 0.003$) but still statistically equal to period 0 of group A1 ($p = 1.0$). For all further measuring periods, the level of perceived stress during corresponding periods is statistically equal ($p = 1.0$ for periods 2-4; $p = 0.145$ for period 5).

2. Differences between different phases of the semester

The level of perceived stress increases throughout the course of both the first and second semesters. Repeated measure ANOVAs show significant variation of the total stress score with similar effect size (first semester, i.e., groups A1 and B1 combined: $F(4.2, 529.1) = 137.1, p < 0.001$, $\eta_P^2 = 0.52$; second semester, i.e., group A2: $F(2.8, 149.4) = 41.1, p < 0.001$, $\eta_P^2 = 0.44$). For the first semester, post-hoc analysis (cf. p-values in Table VI in the appendix) shows that the perceived stress during periods 0 and 1, i.e., between pre-lecture time and the first lecture weeks differs significantly from one another as well as from all further periods in the lecture time ($p < 0.001$). Thereafter, no significant differences between the periods occur. Thus, in the first semester, the level of perceived stress increases during the beginning of the semester but reaches a constant plateau after only three lecture weeks. Similarly, within the second semester, the stress level in period 1 significantly differs from all further measuring periods. Additionally, the stress score in period 5 is significantly higher than in periods 2 and 3 ($p < 0.001$), i.e., unlike for the first semester, the stress perception continues to rise in the last two lecture weeks of the second semester.

B. Workload (RQ2)

Figure 5 shows the weekly study-related workload in hours, according to the students' self-estimation over the semester. Similar to the perceived stress, the three groups report similar workloads in corresponding semester weeks. During the pre-course in the first semester (groups A1 and B1), students spend around 20 h on their studies. Over the first four to six lecture weeks, this workload strongly increases up to a level of around 50 h per week. This level persists until the end of the exam weeks, after which the workload drops back to a level of around 20 h or even lower. During the Christmas break, students report a much lower workload of around 20 h.

1. Differences between the three groups

Comparing the workloads of groups A1 and B1, a mixed ANOVA shows neither interaction effects of temporal evolution and group ($F(4.1, 518.1) = 2.2$, $p = 0.06$, $\eta_P^2 = 0.02$) nor group effects ($F(1, 126) = 2.8$, $p = 0.098$, $\eta_P^2 = 0.02$). Consequently, we again combine the data from groups A1 and B1 when analyzing temporal effects in the next section.

For the matched sample of groups A1 and A2 a repeated measure ANOVA shows a significant temporal effect of the workload ($F(6.3, 241.1) = 108.5$, $p < 0.001$, $\eta_P^2 = 0.74$). A post-hoc analysis reveals that period 1 of group A2 significantly differs from all other measuring periods ($p < 0.02$). In particular, this implies that the average workload in period 1 of group A2 is significantly lower than the workload in the corresponding period of group A1 but still significantly higher than the reported workload in period 0 of group A1. We do not find any significant differences between the reported workload of further corresponding measuring periods within the course of the semester ($p > 0.9$ in all pairwise comparisons).

2. Differences between different phases of the semester

A repeated measure ANOVA of the reported workload data for the first semester, i.e. within the combined groups A1 and B1, reveals significant changes in the workload throughout the lecture weeks ($F(4.1, 521.3) = 405.3$, $p < 0.001$, $\eta_P^2 = 0.76$). Post-hoc pairwise comparisons (cf. Table VI in the appendix) show that the pre-lecture period 0 and the period 1 significantly differ from one another as well as from later periods ($p < 0.001$). Moreover, periods 3 and 4 ($p = 0.002$) and periods 4 and 5 ($p = 0.002$) significantly differ from one another, where the mean reported workload in period 4 is below the reported values in the adjacent measuring periods. All other reported values are statistically comparable.

For the second semester, i.e. group A2, a repeated measure ANOVA reveals a significant change in the reported workload within the lecture time ($F(3.4, 181.6) = 111.7$, $p < .001$, $\eta_P^2 = 0.68$). Post-hoc analysis shows that the workload in period 1 significantly differs from all later measuring periods ($p < 0.001$). Moreover, period 5 significantly differs from both period 2 ($p = 0.024$) and period 3 ($p = 0.030$). None of the further pairwise comparisons gets significant ($p > 0.60$).

3. Correlation between perceived stress and workload

The repeated measure correlation [83] between the total stress score and the reported workload was calculated across all measuring points and all groups, i.e., $N = 3,206$ cases. It is $r_{rm}(2824) = 0.62$, CI 99% $[0.59, 0.65]$, $p < 0.001$.

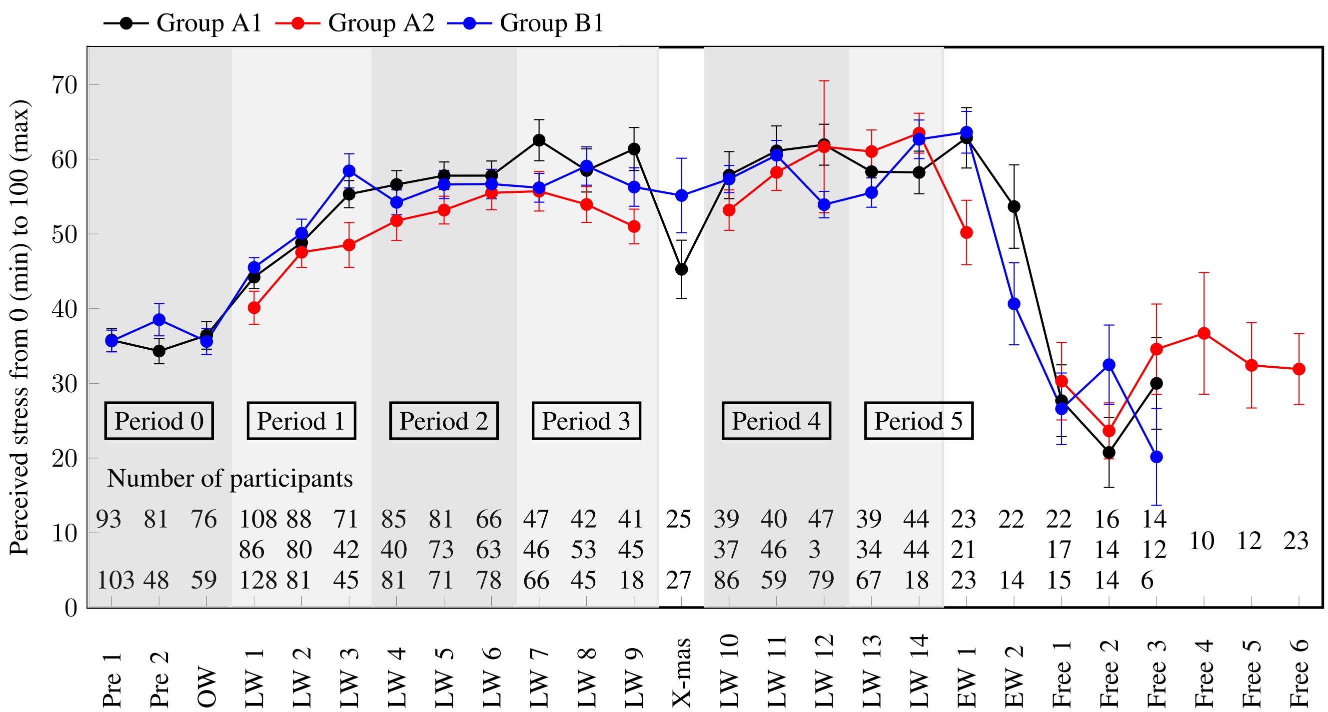


FIG. 4. Perceived stress (mean and standard error per each measuring point) on the artificial scale from 0 (min.) to 100 (max.) of groups A1, A2, and B1, i.e. cohorts A and B in their first or second semester, over the semester with the precourse (Pre), the orientation week (OW), the lecture weeks (LW), the exam weeks (EW), the lecture-free weeks (Free), and the Christmas break (X-mas) in winter semesters. The numbers below indicate the number of participants per measuring point and group (first row for group A1, second for group A2, and third for group B1). The gray areas visualize which measuring points were combined into periods 0 to 5 for quantitative data analysis.

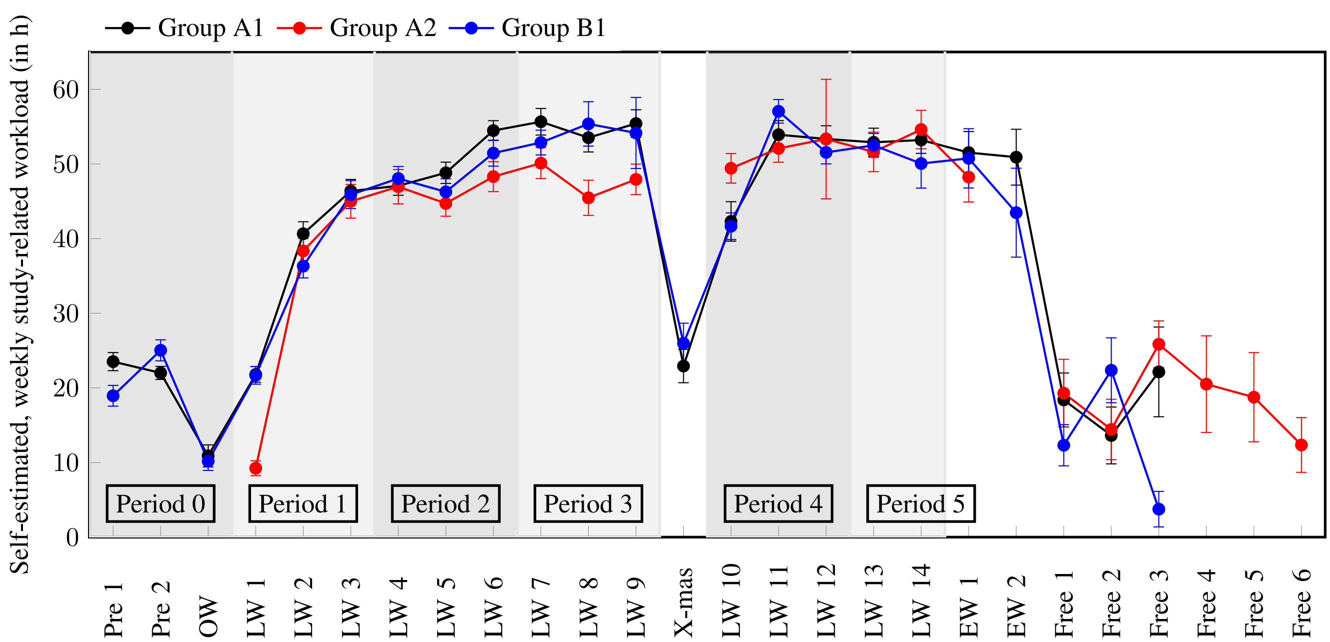


FIG. 5. Workload (mean and standard error per each measuring time) in hours reported by groups A1, A2, and B1, i.e., cohorts A and B in their first or second semester, they have spent in each week over the semester for their studies. The gray areas visualize which measuring points were combined into periods 0 to 5 for quantitative data analysis.

C. Stressors contributing to the stress perception (RQ3)

1. Total frequency of various stressors

Figure 6 presents the percentage of codings per category separately for each group. In all three groups, the category *exercise sheets* (U9) is mentioned most frequently with a share of 15% to 21%. The second and third most common stressors with a share of 9% to 15% are the categories *exams & exam preparation* (U10) and *unspecific mention of courses/subjects* (U11), i.e., a course or subject was mentioned without further information about the underlying stressors. Altogether, these three categories represent 35% to 49% of all responses in the three groups. All further categories were coded less frequently for up to 8% of all responses per group. Every category was used in at least 1% of the codings for one group, 16 out of 20 categories even in at least 3% of the codings for one group.

Table VII in the appendix and, as a summary, Figure 7 provide more details for the frequent categories (U9) and (U11) by further subdividing the codings according to the linked courses. Regarding the *exercise sheets* (U9), 57% of all coded responses do not specify the course associated with the exercise sheet. Out of the remaining responses, 77% pertain to math courses thereby showing significantly less association with physics courses. For groups A1 and A2, the exercise sheets in the course *Mathematics for Physicists* are predominant, accounting for 85% and 67%, respectively, of all responses with identifiable courses. In group B1 the exercise sheets for the courses *Mathematics for Physicists*, *Differential and integral calculus* and *Experimental Physics I* are mentioned with similar frequency, ranging from 31% to 36% of all responses with identifiable courses.

Likewise, the *unspecifically mentioned subjects/courses* (U11) can be further delineated. Across all three groups, 44% of the codings categorized as (U11) are associated with math courses and 32% with the physics lab course, whereas physics and all further courses have a significantly smaller representation. In groups A1 and A2, *Mathematics for Physicists* emerges once more as the most frequently mentioned course (47% resp. 39% of all responses) followed by the physics lab course (22% resp. 33%); all other courses are mentioned in less than 10% of the responses. In group B1, the physics lab course is most frequently listed (45% of all responses), with other courses (e.g., those for higher years of study or in the second subject for teacher training students) occupying the second position (19%). Here, *Mathematics for Physicists* is ranked third, accounting for only 11% of the responses.

In summary, this shows that the math exercise sheets, the math courses itself, in particular *Mathematics for Physicists*, and the physics lab course are the most frequently listed exercise sheets/courses in the codings of (U9) and (U11).

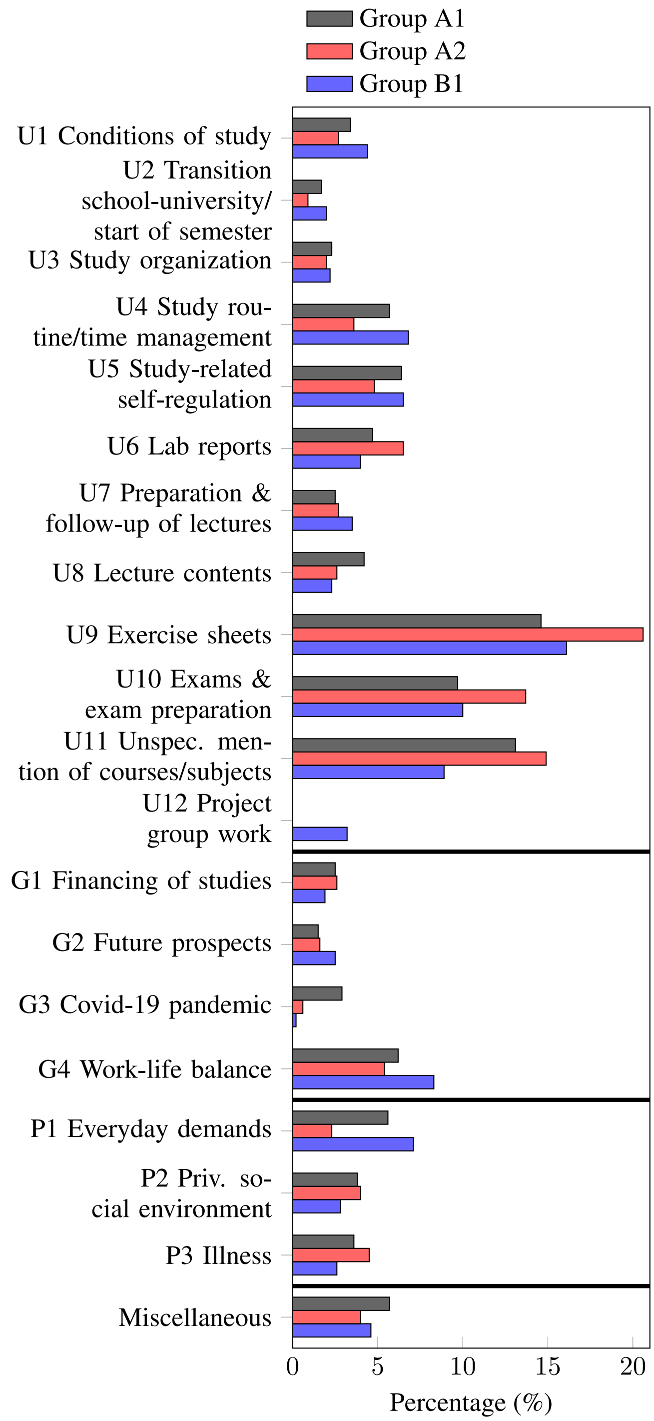
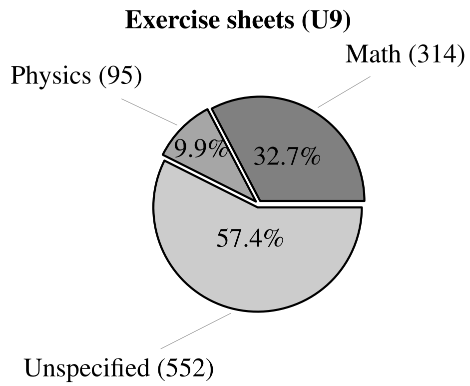


FIG. 6. Percentage of codings per category across all measuring points for groups A1 (2,216 codings), A2 (1,248 codings), and B1 (2,359 codings), i.e., cohorts A and B in their first or second semester.

2. Shift in the frequency of the stressors over the semester

Figure 8 shows how many codings were made in each dimension of the category system for each measuring point for groups A1, A2, and B1. Here, 100% represents the scenario where every participant would have named three stressors each. Consequently, the light gray area illustrates the percentage of open text fields at each measuring point that were left blank by the students. Thus, the figure shows that the relative number of reported stressors varies throughout the semester. On average, students reported 1.7 out of a maximum of three stressors per measuring point. During the 42 LW measuring points, they reported on average $M = 1.9$ ($SD = 0.2$) stres-



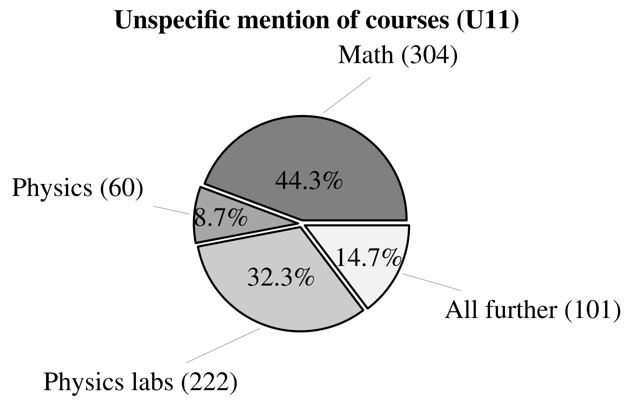


FIG. 7. Percentage of codings within (U9) *Exercise sheets* and (U11) *Unspecific mention of courses* in all three groups that can be linked to math, physics, or lab courses as well as all unspecified or further courses. Absolute numbers are given in brackets.

sors and therefore significantly more than in the 25 other measuring points in other phases of the semester with $M = 1.4$ ($SD = 0.4$) stressors (Welch's t-test, $p < 0.001$, $d = 1.7$). University-related stressors predominate throughout the entire semester, with 60% to more than 90% of all stressors at each measuring point being associated with this dimension. The only exceptions are the pre-course and the orientation week when students were just entering the university and also had to deal with many private and global stressors.

Figure 9 shows the percentage of specific university-related sources of stress mentioned throughout the semester for all three groups. This uncovers distinct phases within the semester in which different stressors become more or less sigificant. Throughout the lecture weeks, the lecture contents and the preparation/follow-up of lectures are frequently mentioned. At the same time, also the exercise sheets are mentioned with a strongly increasing percentage in the first two to three lecture weeks and a nearly steady decline over subsequent weeks. Instead, from the middle of the lecture time, the lab reports and the preparation of exams begin to gain significance. The latter becomes the dominant stressor at the end of the lecture time and during the exam phase. Over the whole lecture time and the lecture-free time thereafter, further courses and subjects are unspecifically mentioned which, as described above, are mostly mentions of the math courses and the lab course during the semester and specific courses (e.g., a programming course) in the lecture-free time.

3. *Correlation between perceived stress and number of stressors*

The repeated measure correlation [83] between the total stress score and the number of mentioned stressors (0 to 3) per participant and measuring point across the total dataset is $r_{rm}(2868) = 0.62$, CI 99% $[0.24, 0.33]$, $p < 0.001$.

V. DISCUSSION

A. Discussion of findings according to RQs

1. *Trajectory of perceived stress (RQ1)*

We can find a characteristic trajectory of perceived stress similar for the first and second semesters and different cohorts of physics students. The similarity of stress perception between the groups A1 and B1 shows that the stress perception in the first semester seems to be comparable on the cohort level. The similarity of the stress perception between groups A1 and A2 might be traced back to similar structures of the first and second semester, e.g., comparable courses and schedules. The significantly lower stress level in period 1 of group A2 could be explained by the fact that most of the students in group A1 had somehow already started with the semester due to the pre-course and orientation week before while the students in group A2 had just started in the new semester during that time as they had a lecture- and largely exam-free time the weeks beforehand.

Regarding the different periods within the semesters, three observations are interesting: First, the pre-course and the orientation week were perceived as less stressful than the lecture weeks which can be explained by the different structures and amount of tasks and demands during these two specific periods within the first semester. Second, the stress perception in the first weeks of the lecture-free time seems to be lower than the initial value in the pre-course for all three groups. This might be explained by a kind of relaxation effect that the students feel relieved after they have written the exams. The slight increase in stress perception in later lecture-free weeks could then be explained with a reduction of this relaxation effect in combination with the second exam phase for some of the students and additional courses like the programming course during that time. Third, the stress perception is lower for group A1 during the Christmas break in comparison to the lecture weeks before and after but not for group B1. This can have various reasons including the fact that primarily students respond to a university-related online survey during vacation time (some students even participated on Christmas Eve) if

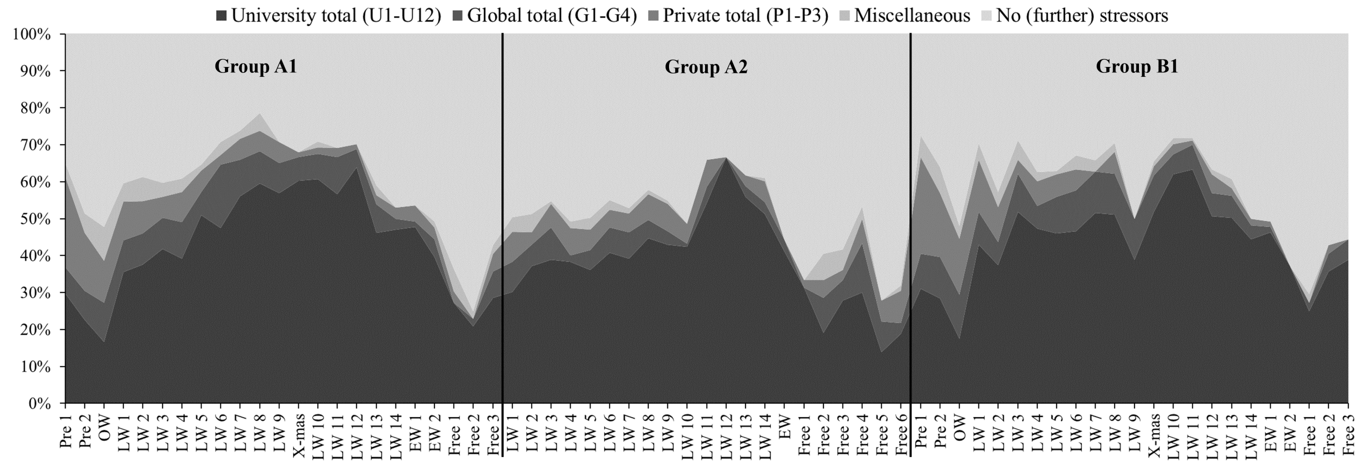


FIG. 8. Percentage of stressors listed by the students per each measuring point for all investigated groups, broken down according to the dimensions of the category system (university, global, private, and miscellaneous, cf. Table III). The light grey area "no (further) stressors" represents the percentage of open text fields that remained empty when a participating student listed no or less than three stressors per measuring point; by this, the graph also comparably visualizes how many stressors were listed per each measuring point.

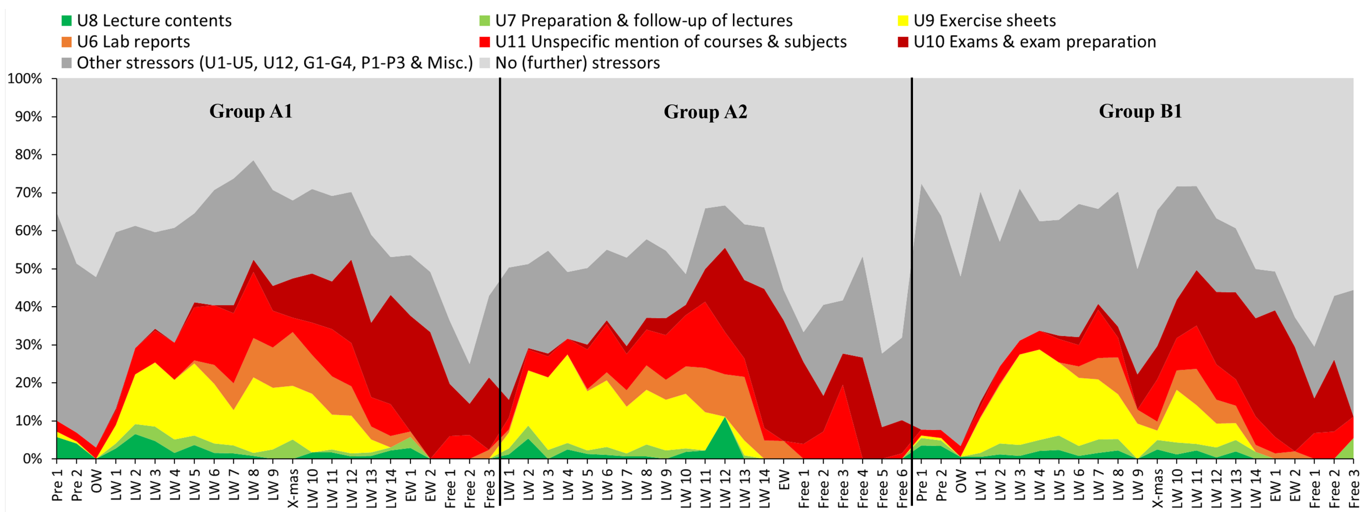


FIG. 9. Percentage of stressors listed by the students per each measuring point for all investigated groups, broken down according to specific categories of the university dimension of the category system (cf. Table III). In dark gray, the codings of all remaining categories are summarized. The light gray area "no (further) stressors" represents the percentage of open text fields that remained empty when a participating student listed no or less than three stressors per measuring point.

they are somehow thinking of the university during that time. Anecdotal reports by some students reveal that stress perception then was caused by the conflict between the desire for vacation and recreation on the one hand and the list of outstanding tasks for their studies in their minds on the other hand.

In Figure 10, we compare selected measuring points from the Göttingen dataset with findings from studies listed in Table I that have also used the PSQ with students. These comparisons need to be done with caution as the cited studies used different versions of the PSQ regarding language, number of items, or rating scale and were conducted in different countries, years, and among varying groups of students. However, the comparison reveals remarkable dynamics in the Göttingen dataset since the measuring points with the minimum and maximum total stress score (Free3 of group B1 and EW1 of group B1) under- and over-exceed all stress scores we can find in the literature and differ significantly from each other (Welch's t-test $t(7) = 6.16$, $p < 0.001$, $d = 3.1$). The score of the focus group, the first measurement Pre1 for groups A1 and B1, is still very low in comparison to the comparative data. The average stress score during LW4 to 14 excluding the Christmas break (i.e., periods 2-5) for groups A1 and B1 (average of the mean stress score of every participating person during the considered measuring points) is much higher than most of all comparative scores in literature, in particular the Göttingen focus group (Welch's t-test $t(416) = 14.94$, $p < 0.001$, $d = 1.4$) and the highest score at the end of the semester measured by Ref. [48] (Welch's t-test $t(284) = 6.48$, $p < 0.001$, $d = 0.3$). The latter is remarkable as the score by Ref. [48] is the best to be compared with our study as the

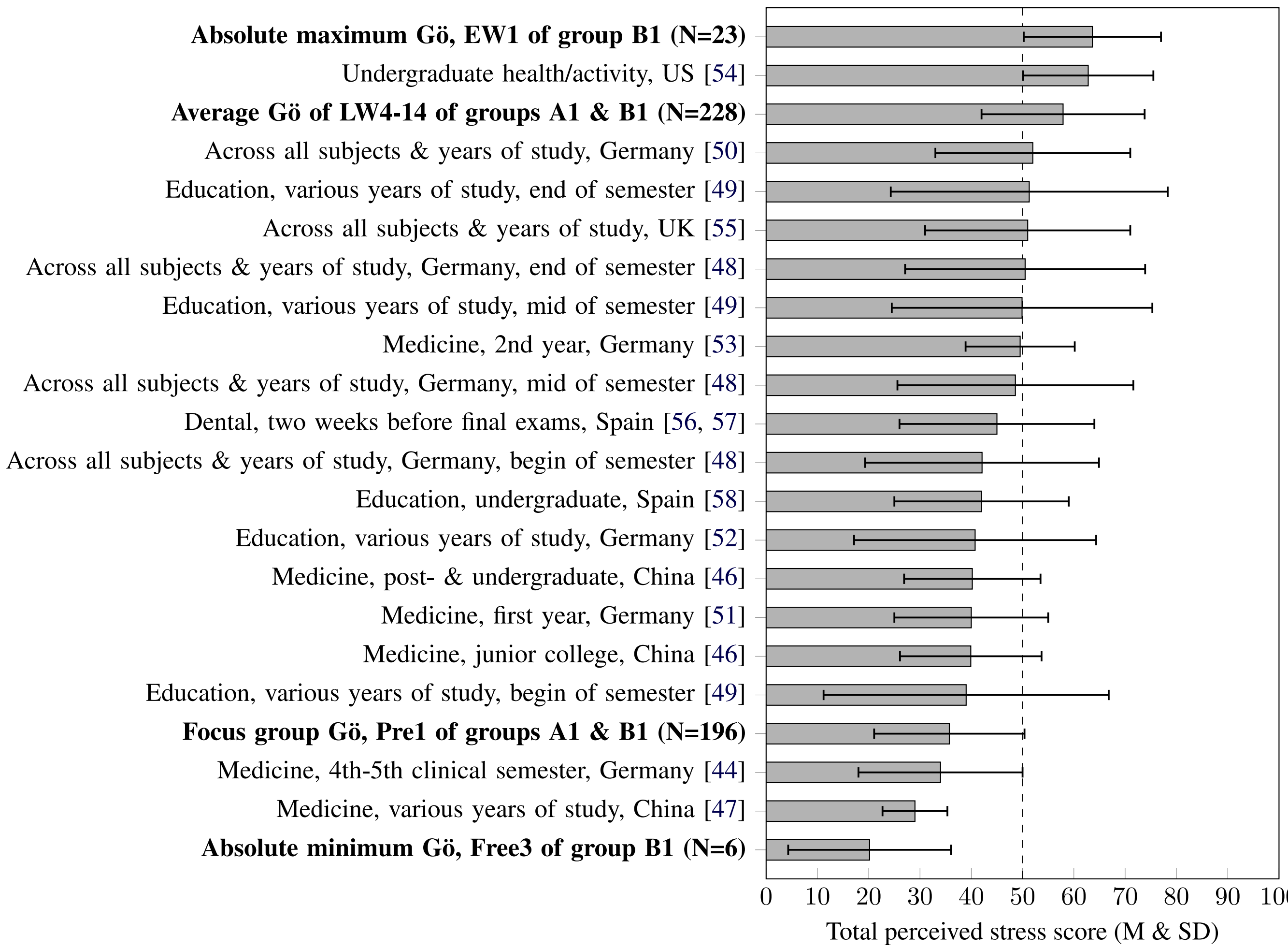


FIG. 10. Comparison of selected total perceived stress scores (mean and standard deviation) from the Göttingen dataset highlighted in bold and with the abbreviation Gö with comparative data from the literature as summarized in Table I.

same PSQ version also with a six-point rating scale was used at a German university in a similar period (summer semester 2020). But still, our findings are in accordance with Ref. [48] regarding that they also identified an increasing stress perception within the first half of the lecture time with no significant changes thereafter until the end of the lecture time.

A similar impression of the physics students' stress levels can be derived by facing the statistically significant differences between the different measuring periods reported in Sec. IV A 1 and by using the scheme for data interpretation from Sec. III B 3. This leads to the observation that the investigated students report, again on average, a *slightly increased* stress perception during most of the lecture time in comparison to pre-course level. This classification needs to be treated with caution, especially for measurements in later lecture weeks, as the students' interpretation of the used items and scale might change over time depending on the current stress perception and underlying stressors.

Nonetheless, the comparison within the Göttingen dataset and with reference data show that physics students feel, on average, stressed during the lecture and exam phase implying that many students even feel very stressed during that time. Such a persisting rather high stress level might be considered es endangering.

2. *Workload (RQ2)*

The quantitative analysis shows that the workload over the semester is equal for different cohorts and similar for the first and second semesters. Only in period 1, group A2 reports a lower workload than group A1 due to the workload of group A2 being significantly lower at measuring point LW1. As for the total stress score, this can again be attributed to the students starting into their second semester with a lower workload as the previous week was still lecture-free time, while groups A1 and B1 already had the pre-course and orientation week before. In the first lecture weeks, the weekly study workload sharply increases leading to a stable plateau of about 50 h during most lecture weeks, 25% higher than a typical full-time job. Some students even report workloads exceeding 70 h in some weeks. During the Christmas break and lecture period 4 thereafter, the reported workload considerably decreased since students naturally spend less time on their studies during Christmas break. Only during the pre-course and the lecture-free time after the exams, a workload of less than 40 h, similar to a full-time job, is reported.

The exact workload measures and the comparison with a full-time job need to be treated with caution as self-estimation of workload based on single-item questions is prone to various distorting effects like social desirability, conformity pressure, and an unconscious orientation at the well-known amount of 40 h for a typical working week [75]. Further, students may estimate their workload differently, e.g., by estimating only the actual time spent on their studies or including the time for breaks, commuting, etc., and, especially during the Christmas break, only those students participate in the survey who are reading their university e-mails during vacation. All these reasons might explain why the workload estimated by the Göttingen students is so much higher than the expected real workload from the time budget study by Ref. [75] even though they have not investigated any physics study program.

Despite the absolute values, the trend in the Göttingen data follows work by Thiel et al. (2006, cited after Ref. [75]) that the daily measured workload in a physics study program is significantly higher during the lecture than the lecture-free time with just a bit more than 20 h/week. However, also in that study, the workload is mostly below the set point of 40 h/week and exceeds this value only in three to four lecture weeks. Average workloads of more than 50 h/week were only reported for two lecture weeks in a chemistry study program, so absolute values in our study are similar for lecture-free time but not for lecture weeks.

All in all, the comparison with reference data suggests that the Göttingen students might have noticeably overestimated their study-related workload but the extent of overestimation is difficult to determine. However, the dynamics within the estimated workload between different phases of the semester are remarkable leading to the conclusion that the workload is not similarly spread over the full semester (lecture vs lecture-free time); resulting peak loads might cause additional stress.

3. Stressors contributing to the stress perception (RQ3)

The analysis of the open text fields reveals that the students' stress perception is caused by several university-related and non-university-related stressors but most of them can be assigned to the university dimension. The predominance of university-related stressors in overall stress perception is probably influenced by the method of data collection, typically conducted within lectures, and the explicitly mentioned context of physics studies. However, the significant impact of these academic stressors on total stress perception should not be overlooked. Strong indicators for this are the fact that the number of listed stressors mentioned per participant during the lecture time is significantly higher than in other phases of the semester and that this number of listed stressors correlates with the total stress score. This is also in line with the findings of Ref. [60] that studying is the area of life that is most frequently perceived by students as a strong stressor.

The category system in Table III, which was inductively derived from the dataset, consists of many categories that can be linked to the state of research [12, 13, 60, 65, 70]. For example, *financing of studies* (G1) including jobbing and *private social environment* (P2), in particular family, can be found in Refs. [13, 60, 65]. *Everyday demands* (P1) are very similar to dimensions like household and housing situation in Ref. [60], *illness* (P3) was also mentioned by Refs.[13, 70]. *Future prospects* (G2) resembles that students often associate stress with fear of the future and uncertainty [60]. *Work-life balance* (G4) picks up the conflicts between the studies and private interests mentioned by Ref. [65] as well as the area of leisure in Ref. [60]. Regarding the university-related dimension, many stressors can be linked to the work by Ref. [70], for example the *labs* (U6), the *exams* (U10), self-study related categories like *preparation and follow-up of lectures* (U7), or the level of difficulty which is part of the categories *lecture contents* (U8) and *exercise sheets* (U9). The importance of the *conditions of study* (U1) was highlighted by Refs.[12, 13, 65], the relevance of time pressure and high quantitative demands, cf. *individual everyday study rountine/time management*, by Refs. [60, 70]. The *study-related self-regulation* (U5) can be compared with the self-concept of ability in Ref. [65].

The findings presented in Figure 6 show that the *exercise sheets* (U9), the *exams and exam preparation* (U10), and the *unspecific mention of courses and subjects* (U11) dominate the university-related dimension of stressors with a share of 9% to 21% of all responses per group over the entire semester. Here, the importance of the two main assessment formats of studying physics becomes prominent, the weekly exercise sheets and the actual exams. The exercise sheets consist of a set of often rather difficult problem-solving tasks that require rather time-consuming work, often in small groups. They are not only important for understanding and practicing the new lecture content knowledge but also for the acquisition of the exam prerequisite. The current data do not explain why categories (U9) and (U11) are mentioned so often. Possible reasons are their exam nature that might be related to a stronger stress perception but also the actual workload since physics students usually spend most time during the semester on the weekly exercise sheets and later the exam preparation. Supplementary, the in-depth analysis shows that primarily the math exercise sheets and correspondingly the math courses and physics lab courses coded in (U11) are more stress-causing than the basic physics courses. This is in line with former research [3, 12, 15, 16, 70] highlighting the difficulty and concurrently relevance of math for study success among science students, as well as the relevance of lab courses for the stress perception [70].

It is noteworthy that the Göttingen students never mentioned the voluntary exercise sheets in the calculus-based course *Mathematical Methods in Physics*, and rarely mentioned the course itself. Instead, they frequently referred to the more algebra- and proof-oriented courses *Mathematics for Physicists* and *Differential and integral calculus*, which have mandatory exercise sheets. This might be explained by the type of course, calculus- versus algebra- and proof-oriented as well as the obligation of the exercise sheets both influencing the perception of the courses. The variations in the frequency of *Mathematics for Physicists* and *Differential and integral calculus* between groups A1 and B1 can thereby be attributed to different course selection behaviors.

It is quite surprising that the topic *financing of studies* (G1) is rather rarely coded, although the financial situation is one of the most important stressors for students according to Refs [60, 65]. A plausible explanation is that while the financing of studies poses a long-term challenge for many students, it may not be perceived as an urgent or acute problem in a specific week compared to other university-related stressors, leading to its less frequent mention in the limited open-text fields of our study. In more detail, Figure 9 reveals that over the semester, stable for both cohorts and similar for winter and summer semesters, specific key stressors can be identified that are more frequently reported. The characteristic progression of the categories *exercise sheets* (U9), *lab reports* (U6), *exams and exam preparation* (U10), and the *unspecific mention of courses and subjects* (U11) demonstrates that in different phases of the semester, different stressors become more or less relevant to the students. This seems to reflect the structure and local conditions of the semester at the University of Göttingen, e.g., that the whole lecture time is accompanied by weekly exercise sheets, the lab course with its experiments and lab reports only starts in the middle of the semester, and the lecture time is followed by the exams which need to be prepared before. The findings are also in line with Ref. [70] who identified the weekly exercise sheets and lab courses as important stressors and a shift of stressors during the semester.

On a meta-level, the existence of characteristic stressors for specific periods of the semester shows that the students have an awareness of their own stress perception and stressors during the semester and that the used open text field question is suitable to reflect this. Thereby, the shifts in the importance of specific categories over the semester must not be over-interpreted because the students were able to state only up to three stressors each, so only the most salient stressors could be acknowledged. The sharp peaks and sinks in some parts of the graph in Figure 9 can be neglected as they are only due to low participation rates at these measuring points.

B. Integrative discussion & possible implications

In this study, we used three different measures, the PSQ, the self-estimated workload, and open text fields for stressors to investigate the first-year physics students' stress perception. The results and discussion show that all three measures give a *coherent* and *complementary* picture of the stress perception. It is *coherent* as the self-estimated workload and the number of listed stressors correlate with the total stress score and somehow depict similar perceptions the students have. The periods within the semester in which students have an increased stress perception coincide with those in which they also report, on average, weekly workloads exceeding those of a typical full-time job. Further, the listed stressors shift over the lecture time while the stress level and workload remain on a stable level. It is also remarkable how similar the three measures are when comparing the three groups, even though the circumstances of each group slightly differ, e.g., by different lecturers, the COVID-19 pandemic for group A1, or the newly introduced project group work for group B1. The varying circumstances across the different groups likely balance out, resulting in a similar level of stress perception at least at the cohort level. The open-text responses further reveal that the three groups report very similar stressors, mainly the exercise sheets, math and lab courses, and exams. These elements might dominate the overall stress perception, leading to comparable total stress scores and workloads across the groups.

The findings are *complementary* as all three utilized measures offer distinct perspectives on the students' stress perception, enriching the understanding of their experiences from multiple angles. The quantitative data from the PSQ and the workload estimations provide insights into the periods of increased stress in the semester and serve as a benchmark for future measurements. The qualitative data helps to explain why the students perceive the physics study entry phase as stressful and therefore serves as a basis for future improvements.

Several implications can be drawn from these observations. Thereby, it is crucial to acknowledge that the high total stress score, substantial workload, or particular stressors cannot and should not be attributed to individual lecturers. All three types of data should be viewed as the outcome of a complex interplay of multiple, interrelated factors that collectively contribute to this perception of stress. The data should also not be interpreted as an argument for blindly simplifying the study programs because students differentiate between stress perceived as stimulating and enriching on the one hand and destructive and unnecessary on the other hand [65]. So, stress and related efforts can also be perceived as challenging and there is also the need for a certain workload and demand to study physics. This highlights the importance of nuanced approaches to addressing stress within academic environments.

Instead, this work can be a starting point for discussing how a study program can be designed that is both appropriately demanding and educating but still not overwhelming for the students. First, it can contribute to increasing awareness of study-related stress and stressors and related aspects like mental health among faculty staff, lecturers, and students. Second, the findings can be used as an empirical basis for the creation and implementation of tailor-made, timed supportive measures as well as the discussion of institutional changes, e.g., as part of reaccreditation procedures. Specifically, the quantitative findings show which phases during the semester are particularly challenging to the students and could therefore be relieved and scaffolded by additional supportive measures. The qualitative findings further reveal that university-related stressors dominate the students' stress perception which can be seen as a good foundation for future supportive measures since they can better be addressed by university faculty than global or particularly private stressors. The codings show what aspects/elements of the study program are central and most stress-causing for the students in different phases of the semester so that the most relevant stressors can be addressed first and that measures can be timed to the phase during the semester where these stressors become dominant. Third, this data can also serve as a benchmark for the comparison with other study programs and universities but also for the evaluation of the effectiveness of newly introduced measures and institutional changes with follow-up measurements.

C. Limitations

Data collection was conducted at only one university. Open text field responses show that some sources of stress and therefore maybe also the stress trajectory depend on local conditions, study structures, and teaching practices. So, the extent to which the stress perception and stressors are university-specific or general remains uncertain. However, the similarity between cohorts A and B suggests that findings might be similar for study programs with a comparable study structure.

Furthermore, the mode of data collection may have influenced the findings: In group A1, we used both paper and online surveys, whereas we only used online surveys later. For as many measuring points as possible, we collected data within the lectures but depending on circumstances in the lectures not always on the very same day of the week or just/additionally online where participation was possible even for a longer period of typically one week from the one to the following measuring point. Both mechanisms led to elongated and sometimes varying periods of data collection. The exact day of participation as well as the question of participation within or outside the lecture might have influenced the instantaneous stress perception and stressors students think of.

As depicted in Figure 4, the participation rate varied from week to week. While, in tendency, more students participated in the first lecture weeks, the number of participants decreased over the semester and was sometimes influenced by specific incidents like canceled lectures (e.g., LW12 of group A2). This trend aligns with the decreasing number of students attending the lectures over the semester as particularly students of higher semesters often skip lectures later in the semester. So, we had the impression that the majority of students who attended on-campus lectures also responded to the survey. Further influences could have been survey participation fatigue after so many measurement repetitions and during lecture-free times as well as study dropout. To encounter the resulting selection bias and gaps in the students' individual trajectories, we used online surveys for a lower participation threshold also for students outside the campus and aggregated measuring periods in the quantitative analysis.

Regarding the open text field questions, workload and stressors, further limitations are the risk of socially desired responses and their very subjective character depending on a high reflection. Especially workload data as responses to single-item questions is known to be prone to various distorting effects [75]. Regarding the stressors, the limited number of three open text fields might have resulted in some students not mentioning certain stressors contributing to perceived stress. There is also a risk that the most immediate or most recent stressors might be more frequently reported by students because of the limitation to three stressors. This recency bias could skew the data in favor of short-term or immediate stressors over long-term or underlying ones and by this, the true diversity or intensity of stressors might be underestimated. The shortness of the text fields sometimes also lead to imprecise and difficult-to-interpret responses. However, facing the correlations between the total stress score and the workload and number of listed stressors per measuring point, and the fact that some students also shared very private, sensible stressors, we consider the mentioned limitations related to open text field questions as not substantial.

D. Outlook: Future perspectives

Our findings and the discussed limitations lead to various future perspectives: First, we would like to extend the investigation of stress perception to other (German) universities to compare different physics study entry phases in the caused stress with each other. This will provide more profound insights into the mechanisms leading to stress and how different circumstances affect stress perception. Moreover, comparative data in higher years of study and different subject domains would improve the understanding of the characteristics of the physics study entry phase.

Second, to encounter the limitations discussed regarding the three open-text fields for the indication of stressors, the categorical system of stressors will be utilized to develop a standardized questionnaire, enabling all students to continuously evaluate the impact of various stressors on their perceived stress on a Likert rating scale. These numeric values will provide a more objective way to compare and contrast stressors and perceived stress levels or to evaluate the efficiency of interventions, and will likely produce more consistent and standardized data in future investigations.

Third, the underlying mechanisms and reasons behind these stressors need to be understood more deeply building a basis for future improvements and supportive measures. Therefore, we will conduct group interviews with both first-year physics students and beyond in which the students reflect on their perceived stressors and already-developed coping strategies and discuss possible improvements that could help to reduce stress perception for future cohorts.

Fourth, we will further investigate relevant individual parameters related to the students' affection and attitude like their sense of belonging to the university (adapted after Ref. [84]) and physics community (adapted after Ref. [85]), the domain-specific growth mindset [86], the motives for choosing a physics study program [13], study-related emotions [87, 88], or self-efficacy, and performance characteristics like the prior math- and physics-related skills at the beginning of their study. This will facilitate a complex, multi-dimensional analysis of relevant factors of the physics study entry phase and stress perception in particular and aims for tailor-made supportive measures.

Fifth, based on all these descriptive-analytical findings, we would like to develop and evaluate concrete supportive measures and study-structural modifications to decrease the students' stress perception in the study entry phase and to cope with the high dropout rate. The very first approach, beyond others, will be the use of a short self-learning intervention supporting a domain-specific growth mindset [89]. Further approaches could be linked to strengthening the students' resilience [90], their sense of belonging, and metacognitive strategies. Refs. [91, 92] have shown that a short two time management training of two to four hours for German undergraduate students can significantly reduce the perceived tension as one of the four subscales of the PSQ and facilitate the perceived control of time. Ref. [50] suggests the implementation of mindfulness-based stress prevention measures. Ref. [54] has shown that for college students „[p]erceived problem solving was a stronger predictor of physical health and perceived stress than were physical activity, alcohol consumption, or social support" [p.360]. This suggests an effective stress management intervention by providing an "academic 'health' course in which the aim is to improve perceived problem-solving abilities, communication skills, and leadership skills for life success, reduced perceived stress, and enhanced physical health" as well as workshops about "improving perceived problem solving skills" and "teaching and practicing skills to be successful and stress free" [p.368]. Moreover, aiming for structural changes, Ref. [65] emphasizes that it is important to identify and address the as unnecessary perceived stressors like a torn day structure because stress is not correlated with satisfaction since students perceive certain stressors also as stimulating and enriching.

Overall, the outlined plans aim for a multiperspective analysis of the physics study entry phase and specific, evaluated support measures tailored to guide the students more effectively through this particularly challenging transition period. The ultimate goals are to lower stress perception, enhance student success, and, hopefully, reduce study dropout rates.

VI. CONCLUSION

In the present study, we investigated the stress perception of first-year physics students at a German university in a three-semester panel study. Stress perception was measured (bi-)weekly for two cohorts of students in their first and one cohort also in the second semester using the PSQ and open text fields for stressors and self-estimated weekly workload. The findings show no major differences between the two cohorts in their first semester or between the first and second semesters of the one cohort that was studied in both semesters. Instead, the stress perception follows a characteristic trajectory with an increasing level in the first lecture weeks of the semester, reaching a stable level of, on average, *slightly increased* stress in the lecture time and exam phase, followed by a decreasing stress level in the lecture-free time afterward. The self-estimated workload shows a similar behavior with a high correlation. The open text fields reveal underlying stressors that are mostly university-related and characteristic of certain phases of the semester. Important stressors are the exercise sheets, especially in math courses, the lab course and lab reports as well as the exam preparation and exams themselves. The study provides a profound and coherent insight into the physics students' stress perception based on quantitative and qualitative measures, which also reveals a high stress perception in comparison to reference data. This can serve as a basis for further studies also at other universities and for prospective supportive measures for the students and institutional changes in physics study programs to improve studyability and reduce high dropout rates.

DATA AVAILABILITY STATEMENT

The full dataset of this study is openly available as supplemental material. There, the open text field responses are coded according to the presented category system as the responses are only in German and very sensitive. They can be shown to interested readers upon request to the authors.

ACKNOWLEDGMENTS

We would like to thank all students in the three investigated courses for their repeated and continuous participation in the survey without which the present paper would have been impossible. We would further like to thank our student assistants Marlene Breither, Stine Gerlach, and Laura Pflügl for supporting the data curation and interrating of the stressors.

ETHICAL STATEMENT

The survey was pseudonymous. Data collection and storage was organized in accordance with the General Data Protection of the European Union and was coordinated with the data protection officer of the University of Göttingen. All participants were informed about the survey goals and the planned data collection and storage beforehand and consented to voluntarily participate. For online surveys, in addition to an access time stamp, only personal data that participants were explicitly asked during the surveys was collected and stored.

AUTHOR CONTRIBUTIONS*

J. O. C.: data curation (supporting), formal analysis (supporting), validation (equal), visualization (supporting), writing - review & editing (equal); **J. H.**: writing - review & editing (supporting); **L. H.**: conceptualization (equal), data curation (supporting), investigation (supporting), methodology (equal), visualization (supporting); **P. K.**: conceptualization (equal), investigation (supporting), methodology (equal), supervision (lead), writing - review & editing (equal), project administration, resources; **S. Z. L.**: conceptualization (equal), data curation (lead), formal analysis (lead), investigation (lead), methodology (equal), validation (equal), visualization (lead), writing - original draft (lead), writing - review & editing (equal); **J. N.**: formal analysis (lead), visualization (supporting), writing - original draft (supporting), writing - review & editing (equal); **S. S.**: supervision (supporting), resources.

* According to CRediT Contributor Roles Taxonomy, https://credit.niso.org

Appendix

TABLE V. Overview of the questionnaire used for each measuring point (demographic data only once at the beginning of each semester).

Questionnaire/Item	Specification/wording (in English)	Specification/wording (in German)
Demographic data	Gender, field of study, current semester, number of university courses currently attended, specification of the courses attended, assessment of own physics performance, assessment of own maths performance, grade of high school diploma	Geschlecht, Studiengang, Fachsemester, Anzahl aktuell besuchter universitärer Veranstaltungen, Spezifizierung der besuchten Veranstaltungen, Selbsteinschätzung der Physikleistung, Selbsteinschätzung der Mathematikleistung, Abiturnote
Perceived Stress Questionnaire (PSQ20)	*In the following you will find a series of statements. Please read each one and rate the frequency with which this statement has been true in your life in the last week (on a scale from 1 = "almost never true" to 6 = "mostly true"). Think about your studies! For each statement, please check the box that best applies. There are no right or wrong answers. Evaluate the answers ad hoc, without thinking too long, and do not leave out any questions.* 20 items, modified after Ref. [44] into a six-point scale (*almost never* (1), *mostly* (6), the gradations in between were not specified)	*Im Folgenden finden Sie eine Reihe von Aussagen. Bitte lesen Sie jede durch und beurteilen Sie die Häufigkeit mit der diese Aussage in der letzten Woche in ihrem Leben zutreffend war (auf einer Skala von 1 = "trifft fast nie zu" bis 6 = "trifft meistens zu"). Denken Sie dabei an Ihr Studium! Kreuzen Sie bitte bei jeder Aussage das Feld an, das am besten zutrifft. Es gibt keine richtigen oder falschen Antworten. Beurteilen Sie die Antworten ad hoc, ohne lange nachzudenken, und lassen Sie keine Frage aus.* 20 Items von Ref. [44], modifiziert durch eine sechsstufige Skala (*fast nie* (1), *meistens* (6), die Zwischenstufen blieben unbenannt)
Self-estimated workload	*Estimate the total amount of time you spent on your studies within the last week (lecture, tutorial, self-study, etc.) in hours.*	*Schätzen Sie den zeitlichen Aufwand, den Sie innerhalb der letzten Woche für Ihr Studium insgesamt aufgewendet haben (Vorlesung, Übung, Selbststudium, etc.) in Stunden.*
Specification of stressors	*Specify up to three causes that are currently generating a load (strongest load first). Be as precise as possible in your indication of what exactly the load consists of.*	*Geben Sie bis zu drei Ursachen an, die gerade eine Belastung erzeugen (die stärkste Belastung zuerst). Seien Sie dabei möglichst präzise in Ihrer Angabe, worin genau die Belastung besteht.*

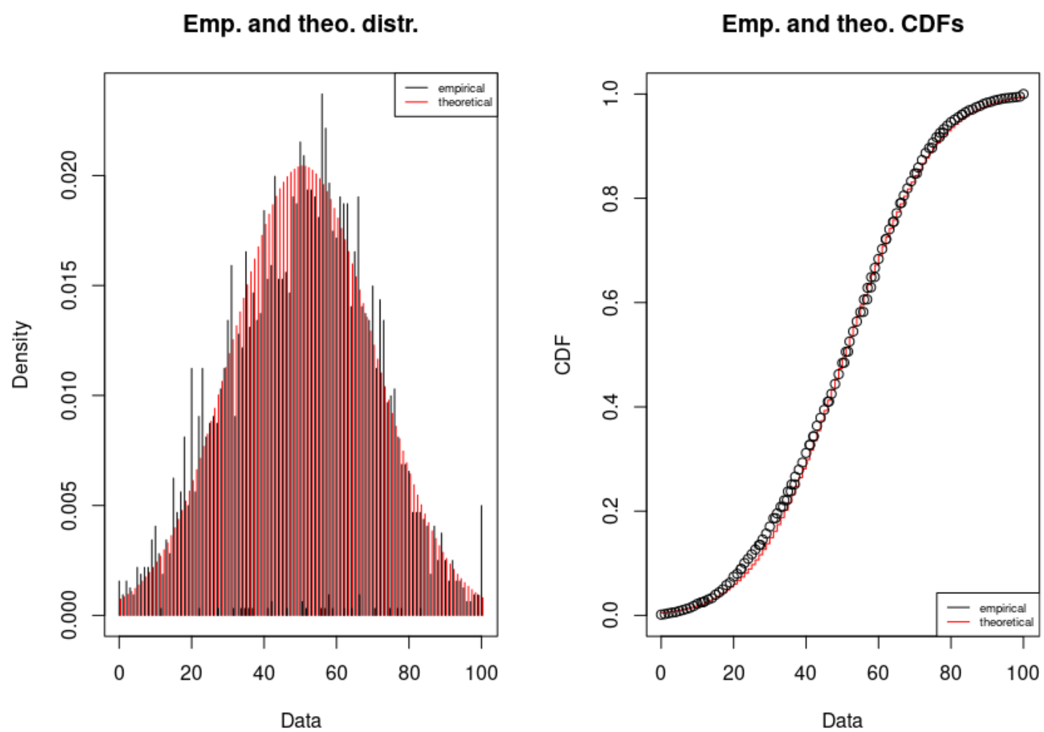


FIG. 11. On the left, the empirical and theoretical density distribution of the total perceived stress scores of the full dataset ($N = 3206$) is displayed. On the right, the according cumulative distribution function (CDF) is displayed. Visual inspection suggests that the total perceived stress scores are nominal distributed.

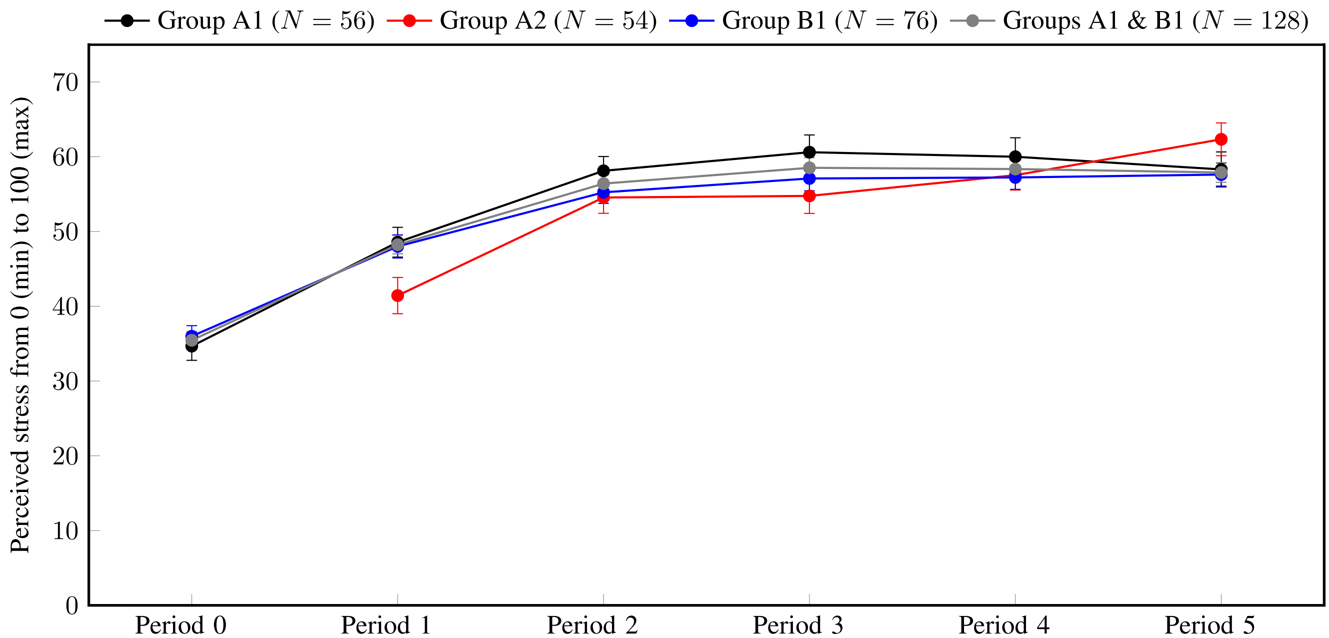


FIG. 12. Visualization of the derived perceived stress scores (means and standard errors) used for quantitative data analysis after selecting specific measuring points (Pre1, Pre 2, OW, and LW1-14), combining them into six measuring periods, and imputing missing data as described in Sec. III C 1. Data is presented for all three groups as well as groups A1 and B1, both first semester, combined as they were treated together when the temporal evaluation of the perceived stress over the semester was analyzed (cf. Sec. IV A 2).

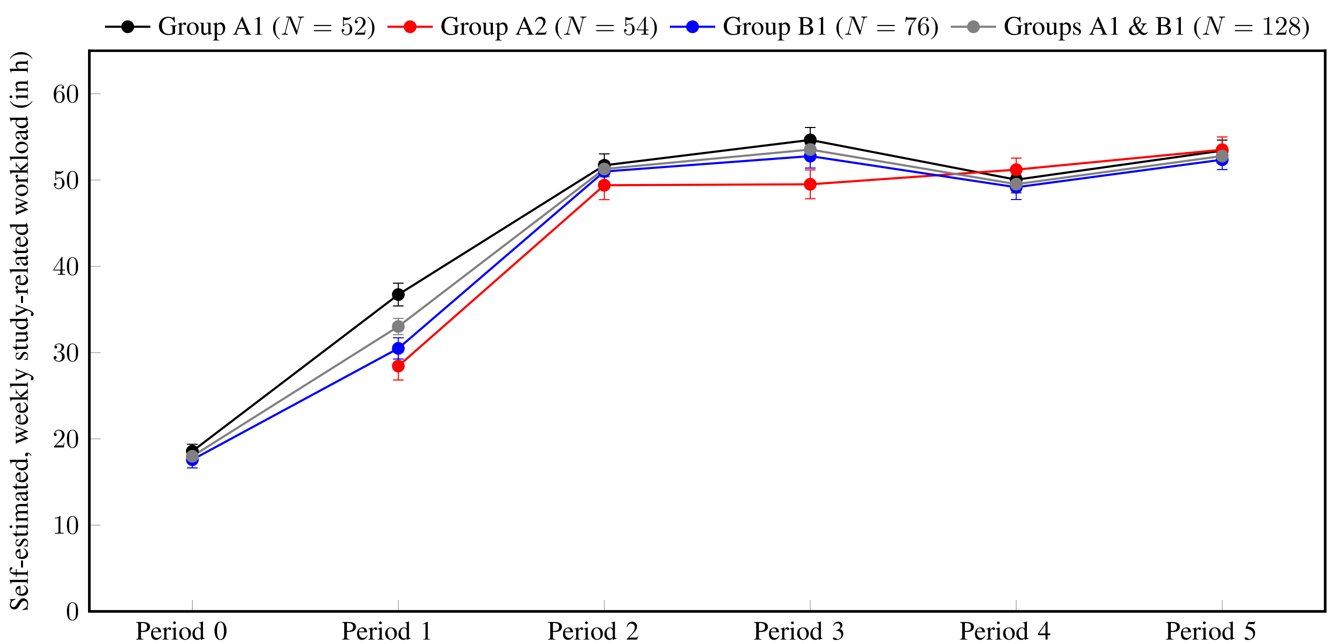


FIG. 13. Visualization of the derived workload (means and standard errors) used for quantitative data analysis after selecting specific measuring points (Pre1, Pre 2, OW, and LW1-14), combining them into six measuring periods, winsorizing, and imputing missing data as described in Sec. III C 1. Data is presented for all three groups as well as groups A1 and B1, both first semester, combined as they were treated together when the temporal evaluation of the perceived stress over the semester was analyzed (cf. Sec. IV B 2).

TABLE VI. Bonferroni-corrected p-Values of the post-hoc analysis of differences of the stress perception and workload between different periods in the semester, separately for the first (groups A1 & B1) and second (group A2) semesters.

	Stress perception (cf. Sec. IV A 2)					Workload (cf. Sec. IV B 2)				
	Period 0	Period 1	Period 2	Period 3	Period 4	Period 0	Period 1	Period 2	Period 3	Period 4
Groups A1 & B1										
Period 1	<0.001					<0.001				
Period 2	<0.001	<0.001				<0.001	<0.001			
Period 3	<0.001	<0.001	0.77			<0.001	<0.001	0.069		
Period 4	<0.001	<0.001	0.92	1.00		<0.001	<0.001	1.00	0.002	
Period 5	<0.001	<0.001	1.00	1.00	1.00	<0.001	<0.001	1.00	1.00	0.002
Group A2										
Period 2		<0.001					<0.001			
Period 3		<0.001	1.00				<0.001	1.00		
Period 4		<0.001	0.25	1.00			<0.001	1.00	1.00	
Period 5		<0.001	<0.001	<0.001	0.052		<0.001	0.024	0.030	0.64

TABLE VII. Percentage (%) of codings within the categories U9 Exercise sheets and U11 Unspecific mention of courses & subjects split up after the different courses in the first and second semester. Data are presented for all three groups A1, A2, and B1, and the total sample. The first row indicates the number of codings for which the percentages in each column are calculated. Gaps indicate that the according course could not be coded for the specific category/group as the course was not offered for that group or had no exercise sheet.

	U9 Exercise sheets				U11 Unspecific mention of courses & subjects			
	Group A1	Group A2	Group B1	Total	Group A1	Group A2	Group B1	Total
N codings / Percentage	167	136	249	552	29	17	41	87
Experimental Physics I/II	4.6	8.9	11.3	8.4	7.9	2.7	3.3	5.1
Mathematics for Physicists I/II	41.4	31.5	10.8	26.6	46.9	38.7	11.4	33.8
Differential and integral calculus I/II	2.5	1.2	12.4	6.0	2.1	5.9	6.6	4.5
Mathematical Methods in Physics	0.0		0.0	0.0	4.5		4.7	3.3
Analytical Mechanics		5.4		1.5		8.6		2.3
Physics lab course					22.1	33.3	44.5	32.0
Precourse					3.4		1.9	2.0
Unspecified	51.5	52.9	65.5	57.4				
Unspecific math course					2.8	0.5	4.3	2.6
Unspecific physics course					0.3	0.0	3.8	1.3
Other courses					10.0	9.1	19.4	12.7

[1] G. Düchs and E. Runge, Studium in der Verlängerung: Statistiken zum Physikstudium in Deutschland 2023, Physik Journal **22**, 33 (2023).

[2] U. Heublein, C. Hutzsch, and R. Schmelzer, Die Entwicklung der Studienabbruchquoten in Deutschland: DZHW Brief 05|2022 (2022).

[3] U. Heublein, J. Ebert, C. Hutzsch, S. Isleib, R. König, J. Richter, and A. Woisch, Zwischen Studienerwartungen und Studienwirksamkeit: Ursachendes Studienabbruchs, beruflicher Verbleib der Studienabbrecherinnen und Studienabbrecher und Entwicklung der Studienabbruchquote an deutschen Hochschulen: Forum Hochschule 1|2017 (2017).

[4] S. Nicholson and P. J. Mulvey, Roster of physics departments with enrollment and degree data, 2021 (2022).

[5] Institute of Physics, Students and graduates in UK physics departments: Data Brief | July 2017 (2017).

[6] K. Klemm, Lehrkräftemangel in den MINT-Fächern: Kein Ende in Sicht: Zur Bedarfs- und Angebotsentwicklung in den allgemeinbildenden Schulen der Sekundarstufen I und II am Beispiel Nordrhein-Westfalens (2020).

[7] OECD, Education at a glance 2020: OECD indicators (2020).

[8] OECD, Education at a glance 2023: OECD indicators (2023).

[9] G. Troendle, Mapping physics students in Europe (2004).

[10] M. Neugebauer, H.-D. Daniel, and A. Wolter, eds., *Studienerfolg und Studienabbruch* (Springer Fachmedien Wiesbaden GmbH, Wiesbaden, 2021).

[11] R. Reissert, Studienabbruch im Widerstreit von Ergebnissen und Meinungen: HIS-Kurzinformation A1/83 (1983).

[12] A. Albrecht, *Längsschnittstudie zur Identifikation von Risikofaktoren für einen erfolgreichen Studieneinstieg in das Fach Physik*, Dissertation, Freie Universität Berlin, Berlin (2011).

[13] F. Thiel, S. Veit, I. Blüthmann, and S. Lepa, Ergebnisse der Befragung der Studierenden in den Bachelorstudiengängen an der Freien Universität Berlin: Sommersemester 2008 (2008).

[14] F. Krause and A. Reiners-Logothetidou, Folgerungen aus dem Studieneingangstest Physik 1978: Vorschläge zum Abbau der festgestellten Defizite, Physikalische Blätter **37**, 295 (1981).

[15] D. Buschhüter, C. Spoden, and A. Borowski, Mathematische Kenntnisse und Fähigkeiten von Physikstudierenden zu Studienbeginn, Zeitschrift für Didaktik der Naturwissenschaften **22**, 61 (2016).

[16] J. Müller, A. Stender, J. Fleischer, A. Borowski, E. Dammann, M. Lang, and H. E. Fischer, Mathematisches Wissen von Studienanfängern und Studienerfolg, Zeitschrift für Didaktik der

Naturwissenschaften **24**, 183 (2018).

[17] S. Sorge, S. Petersen, and K. Neumann, Die Bedeutung der Studierfähigkeit für den Studienerfolg im 1. Semester in Physik, Zeitschrift für Didaktik der Naturwissenschaften **22**, 165 (2016).

[18] D. Woitkowski, *Fachliches Wissen Physik in der Hochschulausbildung: Konzeptualisierung, Messung, Niveaubildung*, Studien zum Physik- und Chemielernen, Vol. 185 (Logos Verlag, Berlin, 2015).

[19] D. Woitkowski and S. Breitkopf, Fähigkeitsselbstkonzept und Lernerfolg im ersten Fachsemester Physik, die hochschullehre **5**, 855 (2019).

[20] D. Woitkowski, L. Rochell, and A. B. Bauer, German university students' views of nature of science in the introductory phase, Physical Review Physics Education Research **17**, 10.1103/PhysRevPhysEducRes.17.010118 (2021).

[21] M. S. Feser, I. Haak, and T. Rabe, Sense of belonging among first-year physics students in germany: Exploring intergroup differences and correlations, EURASIA Journal of Mathematics, Science and Technology Education **19**, em2345 (2023).

[22] E. W. Burkholder, G. Murillo-Gonzalez, and C. Wieman, Importance of math prerequisites for performance in introductory physics, Physical Review Physics Education Research **17**, 10.1103/PhysRevPhysEducRes.17.010108 (2021).

[23] J. D. Edwards, L. Laguerre, R. S. Barthelemy, C. de Grandi, and R. F. Frey, Relating students' social belonging and course performance across multiple assessment types in a calculus-based introductory physics 1 course, Physical Review Physics Education Research **18**, 10.1103/PhysRevPhysEducRes.18.020150 (2022).

[24] Y. Li and C. Singh, Sense of belonging is an important predictor of introductory physics students' academic performance, Physical Review Physics Education Research **19**, 10.1103/PhysRevPhysEducRes.19.020137 (2023).

[25] A. Bhansali and M. D. Sharma, The achievement emotions questionnaire: Validation and implementation for undergraduate physics practicals, International Journal of Innovation in Science and Mathematics Education **27**, 10.30722/IJISME.27.09.003 (2019).

[26] L. Timonen, K. Juuti, and S. Harmoinen, Learning enhancing emotions predict student retention: Multilevel emotions of finnish university physics students in and outside learning situations, Physical Review Physics Education Research **18**, 10.1103/PhysRevPhysEducRes.18.020126 (2022).

[27] I. Haak, *Maßnahmen zur Unterstützung kognitiver und metakognitiver Prozesse in der Studieneingangsphase*, Studien zum Physik- und Chemielernen, Vol. 217 (Logos Verlag, Berlin, 2017).

[28] I. Bausch, R. Biehler, R. Bruder, P. R. Fischer, R. K. Hochmuth, W. Koepf, S. Schreiber, and T. Wassong, eds., *Mathematische Vor- und Brückenkurse: Konzepte, Probleme und Perspektiven* (Springer Fachmedien, Wiesbaden, 2013).

[29] A. Gerdes, S. Halverscheid, and S. Schneider, Teilnahme an mathematischen Vorkursen und langfristiger Studienerfolg. Eine empirische Untersuchung, Journal für Mathematik-Didaktik **43**, 377 (2022).

[30] A. Gerdes, Studienanfängerinnen und -anfänger im Lehramtsstudium Mathematik, ihr Studienverlauf und Studienerfolg, in *Bedarfsgerechte fachmathematische Lehramtsausbildung*, edited by S. Halverscheid, I. Kersten, and B. Schmidt-Thieme (Springer Fachmedien Wiesbaden, Wiesbaden, 2022) pp. 257–274.

[31] C. Goll, Online-Brückenkurs Mathematik des VE&MINT-Projekts (2024).

[32] V. Bach and A. Krieg, Online Mathematik Brückenkurs OMB+ (2024).

[33] A. B. Bauer and P. Reinhold, PSΦ: Entwicklung einer abgestimmten Studieneingangsphase (Physik), in *Lernen, Lehren und Forschen in einer digital geprägten Welt*, edited by H. van Vorst (2023) pp. 539–542.

[34] D. Woitkowski, Komplexitätsgestaffelte Übungsaufgaben zur Unterstützung im ersten Semester Physik, PhyDid B - Didaktik der Physik - Beiträge zur DPG Frühjahrstagung - Bonn 2020 , 85 (2020).

[35] N. Abbas, A. B. Bauer, and P. Reinhold, PSΦ Entwicklung von Unterstützungsmaßnahmen für Theoretische Physik, in *Lernen, Lehren und Forschen in einer digital geprägten Welt*, edited by H. van Vorst (2023) pp. 547–550.

[36] S. Staacks, D. Dorsel, S. Hütz, F. Stallmach, T. Splith, H. Heinke, and C. Stampfer, Collaborative smartphone experiments for large audiences with phyphox, European Journal of Physics **43**, 055702 (2022).

[37] S. Hütz, S. Staacks, C. Stampfer, and H. Heinke, Kleiner Aufwand, großer Nutzen?: Experimentiersets zur Unterstützung experimenteller Übungsaufgaben mit Smartphones, PhyDid B - Didaktik der Physik - Beiträge zur DPG Frühjahrstagung - Aachen 2019 , 273 (2019).

[38] A. Kaps and F. Stallmach, Development and didactic analysis of smartphone-based experimental exercises for the smart physics lab, Physics Education **57**, 045038 (2022).

[39] P. Klein, S. Gröber, J. Kuhn, H. Fouckhardt, G. von Freymann, E. Oesterschulze, A. Widera, A. Fleischhauer, and A. Müller, Teaching experimental physics by using mobile technologies as experimental tools: Videoanalyse-Aufgaben in der Experimentalphysik 1, PhyDid A - Physik und Didaktik in Schule und Hochschule **14**, 1 (2014).

[40] M. A. Ruiz-Primo, D. Briggs, H. Iverson, R. Talbot, and L. A. Shepard, Impact of undergraduate science course innovations on learning, Science **331**, 1269 (2011).

[41] R. S. Lazarus and S. Folkman, *Stress, appraisal, and coping* (Springer Publishing Company, New York, 1984).

[42] R. H. Moos and R. W. Swindle Jr., Person-environment transactions and the stressor-appraisal-coping process, Psychological Inquiry **1**, 30 (1990).

[43] S. Levenstein, C. Prantera, V. Varvo, M. L. Scribano, E. Berto, C. Luzi, and A. Andreoli, Development of the perceived stress questionnaire: A new tool for psychosomatic research, Journal of Psychosomatic Research **37**, 19 (1993).

[44] H. Fliege, M. Rose, P. Arck, S. Levenstein, and B. F. Klapp, Validierung des "Perceived Stress Questionnaire" (PSQ) an einer deutschen Stichprobe, Diagnostica **47**, 142 (2001).

[45] C. Sanz-Carrillo, J. García-Campayo, A. Rubio, M. A. Santed, and M. Montoro, Validation of the spanish version of the perceived stress questionnaire, Journal of Psychosomatic Research **52**, 167 (2002).

[46] R. Meng, J. Li, Z. Wang, Di Zhang, B. Liu, Y. Luo, Y. Hu, and C. Yu, The chinese version of the perceived stress questionnaire: development and validation amongst medical students and workers, Health and Quality of Life Outcomes **18**, 10.1186/s12955-020-01307-1 (2020).

[47] C. Jiang, S. Mastrotheodoros, Y. Zhu, Y. Yang, S. Hallit, B. Zhao, Y. Fan, M. Huang, C. Chen, H. Ma, and R. Meng, The chinese version of the perceived stress questionnaire-13: Psychometric properties and measurement invariance for medical students, Psychology Research and Behavior Management **16**, 71 (2023).

[48] R. Obermeier, M. Gläser-Zikuda, S. Bedenlier, R. Kammerl, B. Kopp, A. Ziegler, and M. Händel, Stress development during emergency remote teaching in higher education, Learning and Individual Differences **98**, 10.1016/j.lindif.2022.102178

(2022).

[49] R. Obermeier, M. Gläser-Zikuda, S. Bedenlier, R. Kammerl, B. Kopp, A. Ziegler, and M. Händel, Wie gestresst sind Lehramtsstudierende im Kontext der Umstellung auf digitale Lehre?, in *Digitale Lehrerbildung*, Lehrerbildung auf dem Prüfstand, edited by U. Stadler-Altmann (Verlag Empirische Pädagogik, Landau, 2021) pp. 152–169.

[50] T. R. Büttner and G. E. Dlugosch, Stress im studium, Prävention und Gesundheitsförderung **8**, 106 (2013).

[51] I. Heinen, M. Bullinger, and R.-D. Kocalevent, Perceived stress in first year medical students - associations with personal resources and emotional distress, BMC Medical Education **17**, 4 (2017).

[52] S. A. Goppert and M. Pfost, Undergraduate students' perceived stress levels in summer term 2020 - a comparison to preceding academic terms, Frontiers in Psychology **12**, 10.3389/fpsyg.2021.672783 (2021).

[53] J.-A. Moldt, T. Festl-Wietek, A. M. Mamlouk, and A. Herrmann-Werner, Assessing medical students' perceived stress levels by comparing a chatbot-based approach to the Perceived Stress Questionnaire (PSQ20) in a mixed-methods study, Digital Health **8**, 1 (2022).

[54] E. Largo-Wight, P. M. Peterson, and W. W. Chen, Perceived problem solving, stress, and health among college students, American Journal of Health Behavior **29**, 360 (2005).

[55] C. M. van der Feltz-Cornelis, D. Varley, V. L. Allgar, and E. de Beurs, Workplace stress, presenteeism, absenteeism, and resilience amongst university staff and students in the COVID-19 lockdown, Frontiers in Psychiatry **11**, 588803 (2020).

[56] J. Montero-Marin, M. M. Piva Demarzo, J. P. Pereira, M. Olea, and J. García-Campayo, Reassessment of the psychometric characteristics and factor structure of the 'perceived stress questionnaire' (PSQ): analysis in a sample of dental students, PLOS ONE **9**, 10.1371/journal.pone.0087071 (2014).

[57] J. Montero-Marín, M. M. Piva Demarzo, L. Stapinski, M. Gili, and J. García-Campayo, Perceived stress latent factors and the burnout subtypes: a structural model in dental students, PLOS ONE **9**, 10.1371/journal.pone.0099765 (2014).

[58] D. Martínez-Rubio, A. Colomer-Carbonell, J. P. Sanabria-Mazo, A. Pérez-Aranda, J. Navarrete, C. Martínez-Brotóns, C. Escamilla, A. Muro, J. Montero-Marín, J. V. Luciano, and A. Feliu-Soler, How mindfulness, self-compassion, and experiential avoidance are related to perceived stress in a sample of university students, PloS one **18**, 10.1371/journal.pone.0280791 (2023).

[59] Y. Luo, B. Gong, R. Meng, X. Cao, S. Tang, H. Fang, X. Zhao, and B. Liu, Validation and application of the chinese version of the perceived stress questionnaire (C-PSQ) in nursing students, PeerJ **6**, 10.7717/peerj.4503 (2018).

[60] A. Ortenburger, Beratung von Bachelorstudierenden in Studium und Alltag: Ergebnisse einer HISBUS-Befragung zu Schwierigkeiten und Problemlagen von Studierenden - Berichtsband - und zur Wahrnehmung, Nutzung und Bewertung von Beratungsangeboten: HIS: Forum Hochschule 3|2013 (2013).

[61] S. Zhao, Y. Zhang, C. Yu, H. Zhang, M. Xie, P. Chen, and D. Lin, Trajectories of perceived stress among students in transition to college: Mindset antecedents and adjustment outcomes, Journal of Youth and Adolescence **52**, 1873 (2023).

[62] U. Herbst, M. Voeth, A. T. Eidhoff, M. Müller, and S. Stief, Studierendenstress in Deutschland: Eine empirische Untersuchung (2016).

[63] J. Brandt and K. W. Kallus, Messung von Stress im Studium, Diagnostica **69**, 111 (2023).

[64] A. Denovan, N. Dagnall, K. Dhingra, and S. Grogan, Evaluating the perceived stress scale among UK university students: implications for stress measurement and management, Studies in Higher Education **44**, 120 (2019).

[65] L. Groß and M.-A. Boger, Subjektives belastungsempfinden von studierenden, in *Die Workload im Bachelor*, edited by R. Schulmeister and C. Metzger (Waxmann, Münster and New York and München and Berlin, 2011) pp. 153–171.

[66] A. Pitt, F. Oprescu, G. Tapia, and M. Gray, An exploratory study of students' weekly stress levels and sources of stress during the semester, Active Learning in Higher Education **19**, 61 (2018).

[67] J. F. Bauer, *Personale Gesundheitsressourcen in Studium und Arbeitsleben: Transaktionales Rahmenmodell und Anwendung auf das Lehramt* (Springer Fachmedien Wiesbaden, Wiesbaden, 2019).

[68] E. Hahn, D. Kuhlee, and R. Porsch, Institutionelle und individuelle Einflussfaktoren des Belastungserlebens von Lehramtsstudierenden in der Corona-Pandemie, in *Das Bildungssystem in Zeiten der Krise. Empirische Befunde, Konsequenzen und Potenziale für das Lehren und Lernen*, edited by C. Reintjes, R. Porsch, and G. Im Brahm (Waxmann, Münster and New York, 2021) pp. 221–238.

[69] C. Vogelsang, Wie belastend war das Corona-Semester?: Belastungserleben Lehramtsstudierender im Praxissemester während COVID-19-bedingter Schulschließungen im Frühjahr 2020, Herausforderung Lehrer*innenbildung – Zeitschrift zur Konzeption, Gestaltung und Diskussion (HLZ) **4**, 265 (2021).

[70] S. Schwedler, Was überfordert Chemiestudierende zu Studienbeginn?, Zeitschrift für Didaktik der Naturwissenschaften **23**, 165 (2017).

[71] S. Cohen, T. Kamarck, and R. Mermelstein, A global measure of perceived stress, Journal of Health and Social Behavior **24**, 385 (1983).

[72] A. Schröder, *Dienstfähigkeit im Lehrerberuf: Ermittlung von Risiko- und Protektivfaktoren in einer vergleichenden Querschnittsstudie dienstfähiger versus dienstunfähiger Lehrerinnen und Lehrer in Hamburg*, Dissertation, Technische Universität Dresden, Dresden (2007).

[73] P. Schulz, W. Schlotz, and P. Becker, *Trierer Inventar zum chronischen Stress*, 1st ed. (Hogrefe, Göttingen, 2004).

[74] B. Gusy, K. Lohmann, and J. Drewes, Burnout bei Studierenden, die einen Bachelor-Abschluss anstreben, Prävention und Gesundheitsförderung **5**, 271 (2010).

[75] R. Schulmeister and C. Metzger, Die Workload im Bachelor: Ein empirisches Forschungsprojekt, in *Die Workload im Bachelor*, edited by R. Schulmeister and C. Metzger (Waxmann, Münster and New York and München and Berlin, 2011) pp. 13–128.

[76] P. Babcock and M. Marks, Leisure college, USA: The decline in student study time: Education outlook, American Enterprise Institute for Public Policy Research (2010).

[77] P. Babcock and M. Marks, The falling time cost of college: Evidence from half a century of time use data, Working Paper Series 10.3386/w15954 (2010).

[78] P. Babcock and M. Marks, The falling time cost of college: Evidence from half a century of time use data, Review of Economics and Statistics **93**, 468 (2011).

[79] M. Händel, M. Stephan, M. Gläser-Zikuda, B. Kopp, S. Bedenlier, and A. Ziegler, Digital readiness and its effects on higher education students' socio-emotional perceptions in the context of the covid-19 pandemic, Journal of Research on Technology in Education **54**, 267 (2020).

[80] J. Bergdahl and M. Bergdahl, Perceived stress in adults: prevalence and association of depression, anxiety and medication in a swedish population, Stress and Health **18**, 235 (2002).

[81] R.-D. Kocalevent, A. Hinz, E. Brähler, and B. F. Klapp,

Regionale und individuelle Faktoren von Stresserleben in Deutschland: Ergebnisse einer repräsentativen Befragung mit dem Perceived Stress Questionnaire (PSQ), Gesundheitswesen (Bundesverband der Arzte des Offentlichen Gesundheitsdienstes (Germany)) **73**, 829 (2011).

[82] J. R. Landis and G. G. Koch, The measurement of observer agreement for categorical data, Biometrics **33**, 159 (1977).

[83] J. Z. Bakdash and L. R. Marusich, Repeated measures correlation, Frontiers in Psychology **8**, 10.3389/fpsyg.2017.00456 (2017).

[84] J. Baumert, W. Blum, M. Brunner, T. Dubberke, A. Jordan, U. Klusmann, S. Krauss, M. Kunter, K. Löwen, M. Neubrand, and Y.-M. Tsai, *Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung von mathematischer Kompetenz (COACTIV): Dokumentation der Erhebungsinstrumente*, Materialien aus der Bildungsforschung, Vol. 83 (Max-Planck-Inst. für Bildungsforschung, Berlin, 2008).

[85] M. S. Feser and T. Plotz, Exploring the validity of a single-item instrument for assessing pre-service primary school teachers' sense of belonging to science, Open Education Studies **5**, 10.1515/edu-2022-0191 (2023).

[86] M. Diederich and V. Spatz, The role of domain-specific growth mindset (implicit theories) at the beginning of STEM university studies: Comparing new and old measures (2024).

[87] R. Pekrun, T. Goetz, and R. P. Perry, Achievement Emotions Questionnaire (AEQ): User's manual: Manual version 2005 (2005).

[88] B. Breyer and M. Bluemke, Deutsche Version der Positive and Negative Affect Schedule PANAS (GESIS Panel) (2016).

[89] M. Diederich, V. Spatz, J. Rehberg, and T. Wilhelm, Weiterentwicklung einer synergetischen Mindset Intervention für die MINT-Studieneingangsphase, in *Lernen, Lehren und Forschen in einer digital geprägten Welt*, edited by H. van Vorst (2023) pp. 551–554.

[90] Y. E. Hofmann, R. Müller-Hotop, M. Högl, D. Datzer, and S. Razinskas, Resilienz stärken: Interventionsmöglichkeiten für Hochschulen zur Förderung der akademischen Resilienz ihrer Studierenden: Ein Leitfaden: IHF Forschungsbericht, 2 (2021).

[91] A. Häfner, A. Stock, and V. Oberst, Decreasing students' stress through time management training: an intervention study, European Journal of Psychology of Education **30**, 81 (2015).

[92] A. Häfner, A. Stock, L. Pinneker, and S. Ströhle, Stress prevention through a time management training intervention: an experimental study, Educational Psychology **34**, 403 (2014).