**Teaching with a Beat: Fitness Trackers as Affordable Solutions for Ubiquitous Teacher Stress Assessment**

**Abstract**

This study investigates the use of wrist-worn fitness trackers to assess teacher stress, focusing on heart rate (HR) as a physiological indicator. Prior research often relied on expensive and invasive methods to measure HR, highlighting the need for affordable, non-invasive alternatives like fitness trackers. We conducted a five-phase lab study with teachers using Fitbit® fitness tracker to monitor HR before, during, and after a teaching session with classroom disruptions. The study examined the correlation between HR data and teachers’ subjective appraisals of stress and their teaching experience. Results showed significant HR increases during teaching sessions, investigating fitness trackers as effective tools for mapping stress in educational settings. As expected, teachers’ HR increased before, peaked during, and progressively decreased after the micro-teaching unit, indicating that wrist-worn fitness trackers are a useful tool to map teachers’ HR before, during, and after teaching. Contrary to expectations, teaching experience and subjective stress appraisals did not significantly predict HR variations. These findings underscore the potential of wearable technology to provide insights into teacher stress and suggest incorporating fitness trackers into teacher training programs to enhance stress management and well-being.

*Keywords:* teacher stress, fitness tracker, heart rate, classroom disruptions, wearable technology, physiological stress measurement

**# Introduction**

In educational contexts, there’s a huge interest in exploring whether heart rate (HR) measures can serve as reliable indicators for teachers’ stress during teaching [@karner2021teachers; @wettstein2020ambulatory]. Prior studies showed that teacher-centered activities and typical classroom-related stressors lead to increased teacher HR in teaching settings [@sperka1995; @scheuch1997psychophysische; @donker2018; @junker2021; @huang2022class]. However, these studies often relied on expensive and invasive electrocardiographs (ECG) to measure teachers’ HR. Therefore, affordable, highly accepted, and non-invasive instruments like wrist-worn fitness trackers [@ferguson2015] could be a valuable tool for analyzing teachers’ HR and the factors contributing to teachers’ physiological stress responses in everyday teaching. Ubiquitous, low-cost assessments of teacher stress would be particularly relevant given the high stress levels in the teaching profession, and the associated negative effects on teachers´ health as well as persistence in the workforce [@johnson2005experience; @montgomery2005meta].

Classroom disruptions are one of the major stressors in teachers’ daily work [@boyle1995structural; @aloe2014multivariate]. According to @lazarus1990theory transactional model of stress and coping, the amount of stress depends on the subjective appraisal of a stressor, which involves considerations about available coping resources. It is, therefore, particularly important for teachers to have sufficient professional and personal resources at their disposal [@cramer2018belastung]. For instance, research has shown that professional knowledge about effective classroom management, including strategies for dealing with classroom disruptions, reduces the risk of teacher stress [@klusmann2012berufliche]. Professional experience is one way how professional knowledge is acquired [@ericsson2006influence]. Experienced teachers typically have more effective classroom management skills for handling classroom disruptions [@wolff2015keeping].

There is a call for research using physiological measures of stress to better understand how stressors like classroom disruptions affect teachers’ stress responses [@wettstein2021]. Teachers’ use of wrist-worn fitness trackers in educational research offers transformative potential by providing detailed in vivo data, allowing researchers as well as teachers themselves to monitor stress during teaching at any time, in any situation, and at low costs. Such ubiquitous, low-cost assessment of stress indicators has the potential to contribute to a better understanding of teacher stress, and eventually to the development of interventions for preventing stress-related, negative consequences for teachers´ health and work. To begin harnessing this potential, this study explored the use of wrist-based fitness trackers as a tool to assess teachers’ HR as an indicator of stress before, during, and after a teaching session in which typical, potentially stressful, classroom disruptions occurred. Teachers’ HR data were triangulated with teachers’ appraisals of classroom disruptions and their teaching experience to establish this approach's validity.

**## Fitness trackers as a ubiquitous, low-cost tool for assessing stress**

Wearables, defined as electronic devices worn directly on the body or integrated into clothing or accessories, serve as versatile data collection solutions [@godfrey2018z] for gathering data like location, movements, and vital signs [@cheng2017underlying]. Fitness trackers, a popular example of wearable technology [@park2020user], provide data on physical activity and cardiovascular parameters such as HR, supporting personalized fitness goals [@nuss2021effects] and stress management [@hao2018chrv]. Their affordability and ease of use have contributed to their widespread use in healthcare, recreation, entertainment, and fitness [sinha2019taxonomy]. Also in education, fitness trackers offer benefits in formal and informal learning environments for both students and teachers [@de2017towards]. Fitness trackers offer ubiquitous, low-cost, and unintrusive data collection [@godfrey2018z], and their use aligns with the increasing popularity and acceptance of wearables among the general population [@peng2022acceptance]. In contrast to self-report questionnaires on stress [@chaplain2008; @liu2020] that are prone to biases like social desirability [@razavi2001self] or recall errors [@van2016accuracy], ambulatory assessment methods [@trull2013ambulatory; @wettstein2020ambulatory] offer objective insights into teachers’ stress levels, e.g., by monitoring teachers’ physiological stress markers without disrupting teaching [@donker2018; @runge2020].

One important health parameter assessed by nearly all wrist-worn fitness trackers is heart rate (HR) [@scalise2018wearables]. HR indicates the number of heartbeats within one minute and is typically expressed as beats per minute (BPM) [@berntson2007cardiovascular; @hottenrott2007]. At rest, the average HR of adults typically ranges between 60 and 80 BPM [@sammito2015guideline]. HR can be detected and measured in different ways using sensors, for example, based on electrocardiogram (ECG) or photoplethysmography (PPG) [@mukhopadhyay2017wearable]. While ECG sensors offer precise measurements by detecting the heart’s electrical activity, their intrusive nature and requirement of direct skin contact may limit their suitability [@kranjec2014non], particularly in educational settings. PPG is a rather uncomplicated and inexpensive technique to measure HR, commonly found in most fitness trackers [@castaneda2018review]. This optical method assesses HR by flashing green or red lights to measure changes in blood volume [@allen2007photoplethysmography].

Physiologically, HR is regulated and influenced on short-time intervals by the sympathetic and the parasympathetic nervous system [@pham2021]. An increase in the activity of the sympathetic system results in HR being speeded up (“fight or flight”) [@taelman2009influence]. In contrast, increased activity of the parasympathetic has the effect of slowing down the HR (“rest and digest”) [@battipaglia2015]. Stress or mental and physical strain are factors that directly influence HR and lead to its increase [@custodis2014heart]. They represent an important physical and emotional stress indicator, as an increased workload is associated with increased HR [@sachs2014]. Therefore, an increase in HR can be regarded as an indicator of increasing stress, and a decrease as an indicator of decreasing stress [@kyriacou1978]. Thus, fitness trackers offer a low-cost, and unobtrusive way of monitoring a wearer’s stress level in many different settings.

**## HR in teaching-learning contexts**

Prior research, not using fitness trackers but traditional electrocardiography (ECG), has shown that changes in teacher HR can be mapped onto stressors experienced by teachers during teaching. For example, teachers’ HR increased during teacher-centered activities when they had to take an exposed position in student-teacher interaction [@sperka1995; @scheuch1997psychophysische; @donker2018; @junker2021]. @sperka1995 for example recorded the HR of 16 pre-service teachers during their first lesson and showed that teachers’ HR increased significantly during teaching. The activation was particularly prominent at the beginning of the lesson and decreased over the course of the lesson. The authors interpreted this finding as indicating that pre-service teachers’ proactive coping strategies, such as actively managing student interactions, helped lower their HR levels. Other ECG studies identified typical stressors predicting increases in HR, such as class size [@huang2022class], or low student engagement and motivation [@junker2021]. For example, @junker2021 recorded the HR of 40 teachers during a real classroom lesson. They provided evidence that teacher stress, induced by factors like low student engagement (e.g., lack of motivation or interest in tasks) or teacher-centered activities (e.g., teacher-focused classroom activities) resulted in elevated HR.

In addition to ECG studies, there are a few studies that used wrist-worn fitness trackers to investigate HR trends in teaching-learning situations [@Darnell2019; @chalmers2021]. @Darnell2019 for example measured the HR of 15 medical college students listening to lecture classes using wrist-worn devices. The analysis revealed a constant decrease in HR from the beginning to the end of the lecture, whereas the HR peak was reached during active learning sessions (peer-discussion based problem solving). @chalmers2021 examined the usability of the average HR, measured with a fitness tracker, to identify physiological changes during stress-inducing tasks (i.e., the Trier Social Stress Test; @kirschbaum1993trier]. The average HR increased significantly from the resting to the stress inducing phases of the task. Even though the participants of both studies [@Darnell2019; @chalmers2021] were not teachers but learners, the results are relevant for studying teacher stress using wearable devices, because the studies showed that a) HR can be effectively recorded using fitness trackers over the course of a whole learning unit, and b) HR changes are in line with the occurrence of activating or stress-inducing tasks.

So far, to the best of our knowledge, only one study has directly assessed teachers’ HR using a wrist-worn fitness tracker during teaching: @runge2020 used a fitness tracker to assess HR as an indicator of stress in *N* = 4 in-service teachers in authentic lessons. They used the fitness trackers’ recordings to create a profile for each teacher, with the aim of differentiating between teachers reporting higher or lower levels of stress. In particular, it was found that the combination of a high HR, a high number of steps, and short sleep duration was characteristic of teachers reporting high stress levels. It should, however, be noted that the generalizability of the results is limited due to the small sample size.

In summary, previous studies have revealed that teachers’ (and students’) HR changes depending on their activities and the stressors they experience, with an increase in HR even before expected stressors occur, and with peaks in activating phases [@Darnell2019; @chalmers2021], whereby teacher-centered phases, in particular, led to an increase in HR [@sperka1995; @scheuch1997psychophysische; @donker2018; @junker2021]. However, studies are still lacking, that investigate data from teacher-worn fitness trackers in larger samples to explore the links between teachers’ HR and subjective stressor appraisal or effects of teaching experience.

**## Emergence of teacher stress and important resources**

The teaching profession is one of the most stressful professions as teachers face a variety of stressors during everyday work [@smith2000; @herman2020; @schult2014belastet]. According to @kyriacou1978, teacher stress can be defined as a negative affective response, typically accompanied by physiological changes such as increased HR, triggered by job-related demands, and perceived as threatening to one’s self-esteem or well-being. Coping mechanisms help to reduce the perceived threat.

This definition of teacher stress is based on the transactional stress model by Lazarus and colleagues [@lazarus1981stressbezogene; @lazarus1984stress], which was modified and tailored to the teaching-learning environment by @kyriacou1978 and later in the German-speaking context by @van2006stress. In general, the transactional stress model [@lazarus1990theory] highlights the interaction between an individual and the environment, whereby stress refers to a person’s subjective reaction to an event (a stressor) that exceeds their adaptive resources.

**Figure 1**

*A model of teacher stress (adapted from van Dick 2006, p.37, modified by the authors)*

Fig. 1 shows, in a simplified way, how classroom events affect teachers’ stress levels, according to the model of teacher stress [@van2006stress]: When potential stressors (e.g., classroom disruptions) occur during teaching (1), teachers intuitively judge how disruptive the event is in a primary appraisal (2). If potential stressors are judged as threatening, i.e., as actual stressors (3), teachers consider whether they have sufficient resources for coping with the stressors (4). Teachers utilize these resources in trying to cope with the stressors, e.g., by employing classroom management strategies (5). In cases where coping fails, stress ensues, often accompanied by physiological reactions like increased HR (6). As part of the coping process, and dependent on its outcomes, teachers re-evaluate the stressor (7).

As shown in Fig. 1, both primary and secondary appraisals are influenced by teachers’ professional experience. As professional experience grows, teachers develop cognitive scripts for managing classroom events, resulting in more complex and effective classroom management skills [@wolff2021classroom]. Effective classroom management strategies in turn are considered to be important resources and have an influence on the stress response. Thus, research has shown that effective classroom management skills and problem-focused coping styles are linked to fewer instances of emotional exhaustion and stress [@maslach2001job; @clunies2008self]. Especially beginning teachers face considerable stress and often feel overwhelmed by the demands of teaching [@ophardt2017klassenmanagement; @wolff2015keeping; @klusmann2012berufliche], with many leaving the profession within the first five years [@ingersoll2003]. Accordingly, when resources are lacking and coping fails, negative consequences for health (e.g., burnout) and for work (e.g., high turnover rates) can arise [@jalongo2006; @unterbrink2007; @aloe2014], highlighting the importance of professional experience in managing stress [@fisher2011].

Despite recognizing these factors, studies often overlook the complexity of the transactional stress model by neglecting the investigation of additional parameters such as professional experience, appraisal, and coping strategies, which are important indicators in predicting stress reactions [for an overview, see @obbarius2021]. @goh2010revised, for example, found direct links in *N* = 129 participants with full-time employment between primary appraisal and initial stress levels, as well as between stress levels over time. In the educational context, @laugaa2008stress showed in a questionnaire study with *N* = 410 French teachers that perceived stress and coping strategies are important variables in explaining variance in burnout. Taken together, professional experience, effective classroom management and active coping styles reduce stress, but more studies combining self-reported data with objective measurements are needed to understand the causes of teacher stress and the interplay of important resources such as professional experience, stress appraisals and stress responses such as an increased HR.

**## Present Study**

The data analyzed in the present study were obtained from in-service and pre-service teachers who participated in a laboratory study as part of a larger project targeting the development of classroom management. As part of the larger project, participants came to the lab individually and taught a 15-minute, self-prepared micro-teaching unit to a “class” of three actors (i.e., trained student assistants) who performed several possibly disruptive, typical classroom events. The micro-teaching unit was potentially stressful for the participants, given its unfamiliar setting and the disruptions of participants’ teaching flow. Thus, in the present study, we were particularly interested in mapping the changes in participants’ HR before, during, and after this micro-teaching unit. We recorded HR data for a total duration of approximately two hours, which were divided in five phases: In the *pre-teaching phase*, participants were welcomed, prepared for the following micro-teaching unit, and familiarized with the setting. In the *teaching phase*, participants taught the micro-teaching unit and experienced possibly disruptive classroom events. In the *post-teaching phase*, participants answered several questionnaires. Next, in the *interview phase,* participants engaged in a stimulated recall interview during which they rated the disruptiveness of each classroom events, and how confident they had felt in dealing with it. In the *end phase*, participants answered another questionnaire. These sequences were identical for all participants. During the entire study, participants wore a fitness tracker on their wrist.

The aims of the present study were twofold:

(1) The first research goal was to investigate whether HR measures assessed by wrist-based fitness trackers are a suitable and effective method for mapping teachers’ HR over the course of the five-phase lab study, including the time before, during, and after the potentially stressful micro-teaching unit.

First, we expected the participants to show an initial increase in their HR, followed by a peak during the micro-teaching unit and a decrease for the remaining phases. In addition, we examined whether z-standardization of the participants’ HR could serve as a useful method to account for individual differences in baseline HR: We expected to observe the same trends in both standardized and non-standardized HR values.

Second, five representative 10-minute intervals were selected from the five phases (see also Figure 2): *pre-teaching interval* (I1), *teaching interval* (I2), *post-teaching interval* (I3), *interview interval* (I4), *end interval* (I5). We examined HR levels and changes during these intervals in order to test the hypotheses that a) teachers showed the highest HR level during the micro-teaching unit (I2), compared to all other phases (\*\*Hypothesis 1a\*\*), and b) that teacher HR increased during the *pre-teaching interval* (I1) while they were preparing for teaching, but decreased in all of the following intervals, because of habituating to (I2) and recovering from (I3-I5) the potentially stressful micro-teaching unit (\*\*Hypothesis 1b\*\*).

(2) We further explored whether teaching experience made a difference in how teachers reacted to the classroom disruptions. In line with the research on teacher expertise and teacher stress reviewed above, we expected more experienced teachers to be less stressed by the classroom events (\*\*Hypothesis 2a). In addition, we were interested in examining the relations between teachers’ appraisals of the classroom events (disruptiveness; confidence in dealing with them) and teachers’ HR level, beyond the explanatory power of teaching experience. We expected higher HR levels for teachers who felt more disrupted, regardless of their teaching experience (\*\*Hypotheses 2b\*\*) and lower HR levels for teachers who felt more confident in dealing with the events, regardless of teaching experience (\*\*Hypothesis 2c\*\*). Lastly, we hypothesized that each of the three predictors (teaching experience, disruption appraisal, confidence appraisal) uniquely contributes to explaining variance in teachers’ HR levels (\*\*Hypothesis 2d\*\*). In addition, we exploratively ran analogous analyses for the *changes* in HR.

**# Method**

**## Participants**

The sample consisted of *N* = 84 pre- and in-service teachers from Germany, who were recruited via personal contact, email lists, and flyers. The data of three participants was lost due to failed data transmission, yielding an analysis sample of *n*total= 81 (*n*total = 52 women, *n*total = 29 men), including 40 pre-service (*n*pre-service = 28 women, *n*pre-service = 12 men) and 41 in-service teachers (*n*in-service = 24 women, *n*in-service = 17 men). Participants had a mean age of 30.95 years (*SD* = 10.90; range: 19-60) and an average teaching experience of 5.64 years (*SD* = 9.46; range: 0-37).

**## Setting and Procedure**

The study was carried out following the ethical standards and the approval of the University’s Institutional Review Board. All participants were informed in detail about the aims of the study before testing. Participation was voluntary and only took place after written consent had been given. Participation was not incentivized.

**Figure 2**

*Procedure of the two-hour study, consisting of five phases with five representative 10-minute intervals as the basis of our analysis.*

**

Each participant came to the lab for a period of approximately two hours in total, and each participant underwent the same phases (see Fig. 2): In the *pre-teaching phase*, the experimenter welcomed the participants and helped them put on the fitness tracker. This was followed by a warm-up session to familiarize the participants with the laboratory setting and the class. This phase took about 10-15 minutes and participants spent this time mostly standing or slowly walking around. During the *teaching phase*, the participants held their self-prepared micro-teaching unit to a class of three trained actors who performed nine, potentially disruptive, classroom events (e.g., chatting with a neighbor, heckling, looking at the phone; see Table ## in the supplementary material for an overview and categorization of all events; and Fig## in the supplementary material for a depiction of the laboratory setting of the micro-teaching unit). In preparation of the micro-teaching unit, the topic and class level could be freely chosen by the teachers with the only requirement that the unit had to be an introductory lesson, and had to consist of supervised individual work and / or frontal teaching. The micro-teaching unit lasted about 15-20 minutes. Participants spent this time mostly standing or slowly walking around. While teaching, participants wore eye-tracking glasses, and their lesson was video-recorded. After having completed the micro-teaching unit, in the *post-teaching phase*, participants filled in questionnaires for approximately 10-15 minutes: a brief computer-based questionnaire assessing sociodemographic data (e.g., teaching experience, gender, studied school type, studied school subjects, extracurricular teaching activities), and a short knowledge test irrelevant to the present study. In the *interview phase*, participants engaged in a Stimulated Recall Interview (SRI). During the SRI, participants watched the video of their own lesson from the ego perspective, recorded through the eye-tracking glasses. The experimenter stopped the video each time one of the nine classroom events happened and asked five open-ended and three rating questions per event. Two of the rating questions are relevant to the present study: the disruption and the confidence appraisal ratings (see Measures). The interview lasted about 45-60 minutes and participants’ position was seated. The *end phase* lasted about 10-15 minutes and participants answered another questionnaire irrelevant to the present study, again in a seated position.

**## Measures**

**### Heart Rate Data and Heart Rate Intervals**

To measure teachers’ HR, we used the wrist-based fitness tracker Fitbit® Charge 4. In line with the manufacturer's instructions [@fitbitnd], the device was attached to the participants’ nondominant hand, a finger’s width above the wrist bone. The tracker works by flashing green LEDs hundreds of times per second, using light-sensitive photodiodes to catch the reflected light, to calculate the volume changes in the capillaries. From this, the tracker calculated how many times the heart beats per minute. HR measurements were generated at least every 15 seconds[[1]](#footnote-1). The raw data contained the estimated HR in BPM for each time stamp. To account for individual differences in the baseline HR, we also calculated z-standardized HR values, e.g., the individual mean value from the unstandardized mean HRs over the duration of the study at the subject level of *n* = 81 participants (i.e., mean standardized HR).

Since we aimed to explore teachers’ HR over the course of the above-mentioned five phases and to keep measurement intervals comparable within phases, we aggregated HR over a representative 10-minute interval within each phase. Previous research has indicated that 10-minute intervals are a useful duration for analyzing PPG data [@lu2008can]. The intervals were selected based on the following rules: The *pre-teaching interval* (I1) comprised the first 10 minutes after the fitness tracker had been put on. The *teaching interval* (I2) started two minutes after the teacher had started the teaching unit. This interval was of the highest relevance to our study. We explicitly chose an early 10-minute interval within the *teaching phase*, as previous studies revealed that the beginning of a lesson is essential and demanding regarding teacher-student interaction [@donker2018quantitative; @claessens2017positive]. The *post-teaching interval* (I3) started immediately after the end of the teaching unit. The *interview interval* (I4) was defined as the mid-10 minutes between the end of the teaching unit and the time point when the fitness tracker was taken off so that all participants were being interviewed during this interval. The *end interval* (I5) comprised the last 10 minutes before the fitness tracker was taken off.

**### Teaching Experience**

The participants’ teaching experience was assessed as a part of sociodemographic data. Participants stated their work experience in years (excluding the traineeship year that is common in Germany).

**### Subjective appraisal of the classroom events and coping processes**

The subjective disruption and confidence appraisals were assessed during the SRI on an 11-point rating scale, ranging from 0 (not at all disrupting/confident) to 10 (extremely disrupting/confident). For the current analysis, ratings were averaged across the nine classroom events for each participant, as we were interested in the general stressfulness of the *teaching phase* for each participant, specifically, in the aggregated effect of all potentially stressful events (disruption rating) and the mean level of subjective coping (confidence rating).

**## Data analysis**

We conducted all analyses with R [@RStudio2020]. Graphics were created using ggplot2 (v3.3.3; Wickham, 2016).

\*\*Research goal 1\*\*. The first research goal included mapping teachers’ HR before, during, and after the micro-teaching unit over the course of the five-phase lab study.

Regarding the teachers’ HR trend, we displayed the HR trend over the course of the entire study. We visually compared unstandardized and standardized HR trends over the course of the entire two-hour study.[[2]](#footnote-2) For all further analyses, we used standardized rather than unstandardized HR values.

For testing Hypothesis 1a, which examined mean differences of the HR levels across the selected five intervals, we initially conducted a one-way ANOVA with repeated measures as an omnibus test. The dependent variable was the mean standardized HR, which was calculated by averaging the mean standardized HR for each interval for each person, resulting in 81 values. To identify the interval with the highest HR level (i.e., mean standardized HR), we subsequently conducted *t*-tests with planned contrasts as post-hoc tests, as well as the effect size *d* [@cohen1988new]. Specifically, we tested the mean differences between the *teaching interval* (I2) and the other four intervals (see Table 1).

For testing Hypothesis 1b, which examined the HR changes (i.e., mean slopes) within each interval, we conducted a linear estimation of the increase or decrease in standardized HR values over time. To this end, we used fixed intercept fixed slope regression models [@gelman2006data] for each interval to estimate intercepts and linear slopes for all individuals which were then averaged across individuals. Mean slope[[3]](#footnote-3) and mean intercept estimates were based on all values at all measurement points per interval for all participants (see Table 2). Mean slope and mean intercept values represent the unstandardized coefficients.

\*\*Research goal 2\*\*. Addressing our second research goal, we examined the effects of teaching experience and subjective appraisal of disruptive classroom events on teachers’ HR levels (i.e., mean standardized HR) during the five phases.

To test Hypothesis 2a, we examined the effect of teaching experience on participants’ HR levels (i.e., mean standardized HR)[[4]](#footnote-4) for each of the five intervals using linear regression models with teaching experience as the sole predictor. To test Hypotheses 2b and 2c, we separately augmented the model by either teachers’ disruption appraisal (Hypothesis 2b) or confidence appraisal (Hypothesis 2c) as predictors, while controlling for the shared variance with teaching experience. To test Hypothesis 2d, we examined the effects of all three predictors in one regression model. Furthermore, we repeated these steps to explore the effects of teaching experience and subjective appraisals on *changes* in teachers’ HR (i.e., mean slopes) at each interval. Please note: HR levels and changes were not regressed on the disruption and confidence appraisals in the *pre-teaching interval* (I1), because the appraised classroom events had not yet taken place in the *pre-teaching interval*.

**# Results**

**## Research goal 1: Mapping teachers’ HR over the course of the study phases**

The first part of our first research goal was to map participants’ overall HR trend and explore whether z-standardization of participants’ HR is a useful method to account for individual differences in the baseline HR. Means, standard deviations, and range of teachers’ unstandardized and standardized HR are shown in Table 1. Fig. 3 a. and b. display the unstandardized and standardized HR trends, respectively.  HR initially increased, peaked, and then decreased, with the unstandardized and standardized HR graphs showing high similarity. Thus, for all further analyses, we used participants’ standardized HR values.

**Table 1**

*Mean HR (*M*), standard deviations HR (*SD*), and range of teachers’ HR over the course of the entire study and the five intervals (unstandardized in BPM/z-standardized)*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Interval | *M HR* | *SD HR* | Min | Max | |
| Overall Course of 2h | 90.09/0.041 | 15.76/0.991 | 512/-4.03 | 164/4.56 | |
| Pre-teaching interval (I1) | 96.28/0.48 | 14.11/0.88 | 56/-3.56 | 139/3.24 | |
| Teaching interval (I2) | 100.80/0.85 | 16.23/0.77 | 63/-2.18 | 164/4.37 | |
| Post-teaching interval (I3) | 93.61/0.27 | 14.01/0.76 | 60/-2.17 | 150/3.06 | |
| Interview interval (I4) | 82.32/-0.72 | 11.85/0.74 | 51/-2.51 | 132/4.39 | |
| End interval (I5) | 77.95/-1.07 | 11.14/0.57 | 502/-2.68 | 120/2.96 | |
| 1 Please note that *M* and *SD* of the overall course vs. the individual intervals were not exactly 0 and 1 due to rounding differences.  2 Deviations of the minimum values in the overall course vs. the *end interval* (I5) are due to data of a few participants who needed more than two hours to finish the study. | | | | |

**Figure 3**

*Overall course of the HR with the unstandardized HR in BPM shown in Fig. 3a. and the z-standardized HR shown in Fig. 3b. for the planned 2-hour study*



We first tested the hypothesis that teachers showed the highest mean standardized HR during the micro-teaching unit, compared to all other phases (Hypothesis 1a). Repeated measures ANOVA revealed that the mean standardized HR differed statistically significantly between intervals, *F*(4, 400) = 260.62, *p* < .05, *f* = 1.60 (large effect). Post-hoc contrasts indicated that, as hypothesized, the mean standardized HR was significantly higher in the *teaching interval* (I2) than in all other intervals (see also Fig. 4). Specifically, it was higher than in the *pre-teaching interval* (I1; *t*(400) = -10.08, *p* < .05, *d* = 1.03; large effect), the *post-teaching interval* (I3; *t*(400) = -6.94, *p* < .05, *d* = 1.37; large effect), the *interview interval* (I4; *t*(400) = 15.00, *p* < .05, *d* = 3.29; large effect), and the *end interval* (I5); *t*(400) = 22.54, *p* < .05, *d* = 4.64; large effect).

**Figure 4**

*Distribution of the mean standardized heart rate in the five intervals*

**

Next, we examined HR changes (i.e., mean slopes) within each interval to test for the hypothesis that HR would increase in the *pre-teaching phase* and decrease in all other phases (Hypothesis 1b). The mean intercepts and mean slopes, complemented by their standard deviations for each interval, are shown in Table 2; the graphical representation of the slopes is displayed in Figure 5. The mean slope of the *pre-teaching interval* (I1) was significantly positive, indicating an increase in HR, as hypothesized. Further, the mean slopes of the *teaching interval* (I2) and *post-teaching interval* (I3) were significantly negative, indicating a decrease in HR. For the last two intervals, the *interview interval* (I4) and *end interval* (I5), the mean slope was negative but did not differ significantly from zero.

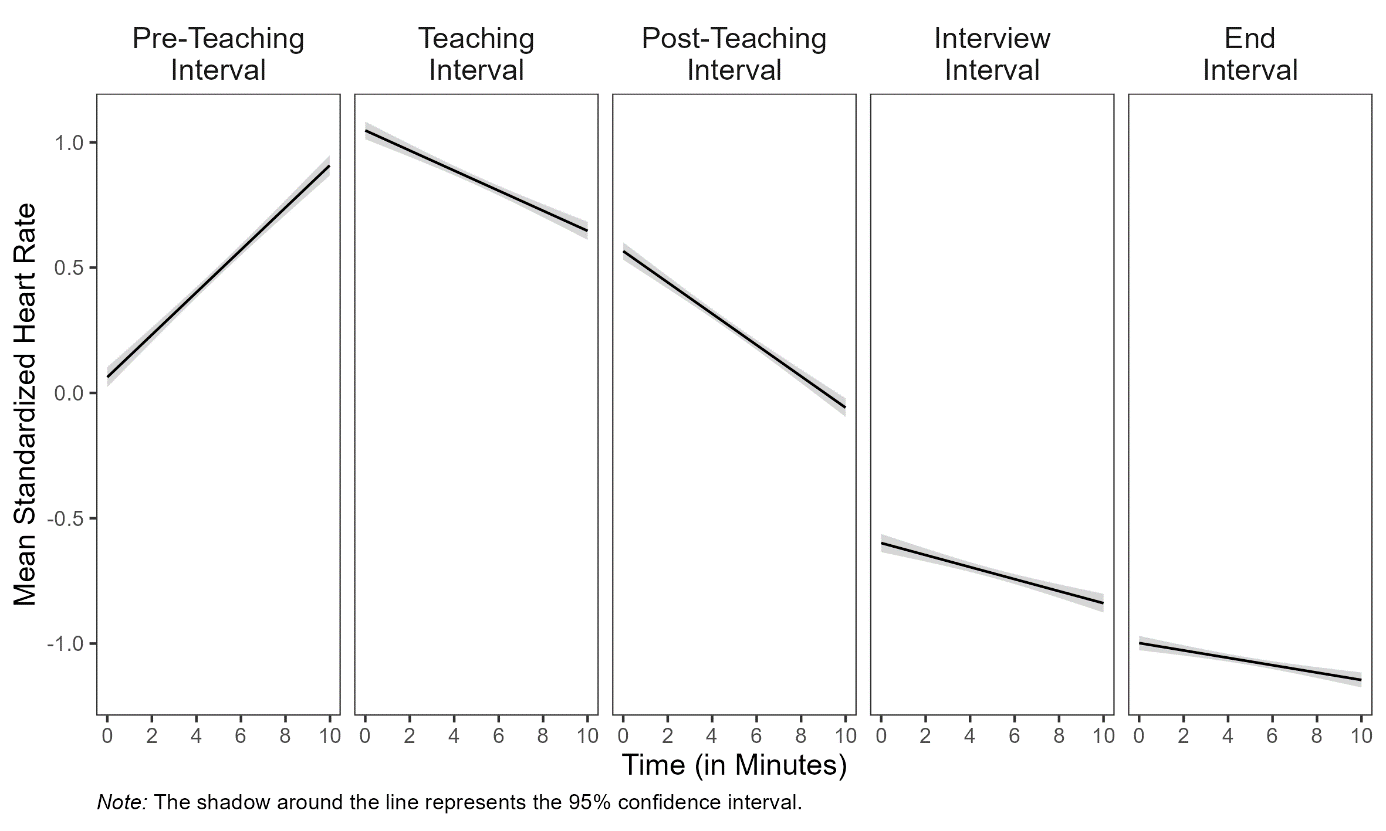
**Table 2**

*Descriptive statistics* *(*n, M, SD*)* *for the mean intercepts and the mean slopes for the five intervals*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Interval | n1 | *M (SD)* | | *p* | |
|  |  | Intercept | Slope | Intercept | Slope |
| (1) Pre-teaching interval | 6896 | 0.052 (0.820) | 0.085\* (0.133) | .57 | < .05 |
| (2) Teaching interval | 7150 | 1.025\* (0.690) | -0.039\* (0.108) | < .05 | < .05 |
| (3) Post-teaching interval | 6664 | 0.549\* (0.547) | -0.060\* (0.101) | < .05 | < .05 |
| (4) Interview interval | 6287 | -0.617\* (0.614) | -0.022 (0.070) | < .05 | .006 |
| (5) End interval | 5990 | -1.004\* (0.500) | -0.012 (0.074) | < .05 | .14 |
| *Note.* \* *p* < .05  1All measurement points per interval for all participants. Note that the variation in *n* stem from the variation in the number of collected data points by the fitness tracker. | | | | | |

**Figure 5**

*Graphical display of the linear trend of the mean standardized HR for the five intervals*



**## Research goal 2: Prediction of mean standardized HR and mean slopes with teaching experience and subjective appraisals**

Correlations among mean standardized HR/mean slopes, teaching experience, disruption appraisal, and confidence appraisal are presented in Table 3. Correlations between mean standardized HR/mean slopes and the other variables were mostly very small and statistically non-significant, except for the *pre-teaching interval* (I1), in which mean slope and teaching experience correlated negatively (*r*=-.27), and the *interview interval* (I4) in which mean HR and teaching experience correlated positively (*r*=.24). Correlations between teaching experience and appraisals were substantial: more experienced teachers had lower disruption appraisals (*r*=-.36), and higher confidence appraisals (*r*=.44). Moreover, the two appraisal variables were negatively correlated (*r*=-.37).

Concerning the effect of teaching experience on participants’ HR levels (i.e., mean standardized HR) for each of the five intervals (testing Hypotheses 2a-d), teaching experience significantly predicted mean standardized HR only in the *interview interval* (Table 4, Interview interval, Model 1), indicating a higher mean standardized HR for teachers with more teaching experience. This relationship is, in fact, in the opposite direction as predicted by Hypothesis 2a.

Adding neither disruption appraisal (\*\*Hypothesis 2b\*\*) nor confidence appraisal (\*\*Hypothesis 2c\*\*) while controlling for the shared variance with teaching experience revealed any significant effects on teachers’ mean standardized HR.

When considering the effects of the three predictors in concert (\*\*Hypothesis 2d\*\*), mean standardized HR was significantly predicted only by disruption appraisal, and only in the *post-teaching interval* (Table 4, Post-teaching interval, Model 4), indicating a higher mean standardized HR for teachers who felt more disrupted by the classroom events, when controlling for all other factors.

Concerning the explorative investigation of the effects of teaching experience and subjective appraisals on *changes* (i.e., mean slopes) in teachers’ HR, teaching experience significantly predicted the mean slope in the *pre-teaching interval* (Table 4, Pre-teaching interval, Model 1), indicating a less steep HR increase in teachers with more teaching experience. For all other intervals, the prediction was not significant.

**Table 3**

*Correlations between mean standardized HR/mean slopes and the predictor variables of teaching experience (TE), disruption appraisal (DA), and confidence appraisal (CA) for the five intervals*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | Pre-teaching interval | Teaching  interval | Post-teaching  interval | Interview  interval | End  interval |
| Teaching Experience | − .17/− .27\* | .11/−.02 | − .04/−.03 | .24\*/−.20 | .04/.11 |
| Disruption Appraisal | − .01/.16 | − .20/.08 | .20/−.14 | − .13/.01 | .04/.12 |
| Confidence Appraisal | − .10/− .18 | .06/.09 | .04/−.03 | .09/−.19 | − .07/.13 |
| *Note.* \* *p* < .05. | | | | | |

**Table 4**

*Standardized regression coefficients of mean standardized heart rate and mean slopes predicted by teaching experience, disruption appraisal, and confidence appraisal for the five intervals*

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | *Dependent variable: mean standardized HR and mean slopes* | | | | | | | | | | | | | | | |
|  | Model 1 | | | | Model 2 | | | | Model 3 | | | | | Model 4 | | | |
|  | Std. mean HR | | Mean slopes | | Std. mean HR | | Mean slopes | | Std. mean HR | | Mean slopes | | | Std. mean HR | | Mean slopes | |
|  | β (SE) | *p* | β (SE) | *p* | β (SE) | *p* | β (SE) | *p* | β (SE) | *p* | | β (SE) | *p* | β (SE) | *p* | β (SE) | *p* |
| **Pre-teaching interval (I1)1** |  |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |  |
| Teaching  Experience | -.17  (.005) | .12 | -.27\*  (.002) | <.05 |  |  |  |  |  |  | |  |  |  |  |  |  |
| R2 | .030 |  | .071 |  |  |  |  |  |  |  | |  |  |  |  |  |  |
| **Teaching interval (I2)** |  |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |  |
| Teaching  Experience | .11  (.002) | .34 | -.02  (.001) | .83 | .04  (.005) | .73 | .01  (.001) | .96 | .10  (.006) | .42 | | -.08  (.001) | .54 | .05  (.006) | .67 | -.05  (.001) | .72 |
| Disruption  Appraisal |  |  |  |  | -.18  (.041) | .13 | .08  (.010) | .50 |  |  | |  |  | -.19  (.042) | .13 | .12  (.010) | .34 |
| Confidence  Appraisal |  |  |  |  |  |  |  |  | .01  (.046) | .92 | | .12  (.011) | .34 | -.04  (.047) | .76 | .15  (.012) | .24 |
| R² | .012 |  | .000 |  | .040 |  | .015 |  | .012 |  | | .010 |  | .042 |  | .031 |  |
| ∆ R² |  |  |  |  | .028 |  | .015 |  | .000 |  | | .010 |  | .030 |  | .031 |  |
| **Post-teaching interval (I3)** |  |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |  |
| Teaching  Experience | -.04  (.005) | .70 | -.03  (.001) | .80 | .04  (.005) | .76 | -.09  (.001) | .44 | -.08  (.006) | .55 | | -.02  (.001) | .89 | -.01  (.006) | .91 | -.07  (.001) | .61 |
| Disruption  Appraisal |  |  |  |  | .22  (.040) | .07 | -.18  (.009) | .14 |  |  | |  |  | .25\*  (.041) | <.05 | -.20  (.010) | .12 |
| Confidence  Appraisal |  |  |  |  |  |  |  |  | .08  (.045) | .55 | | -.03  (.011) | .83 | .14  (.046) | .27 | -.08  (.011) | .54 |
| R2 | .002 |  | .001 |  | .043 |  | .020 |  | .006 |  | | .002 |  | .058 |  | .023 |  |
| ∆ R2 |  |  |  |  | .041 |  | .019 |  | .004 |  | | .001 |  | .056 |  | .022 |  |
| **Interview interval (I4)** |  |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |  |
| Teaching  Experience | .24\*  (.006) | <.05 | -.20  (.001) | .07 | .22  (.006) | .06 | -.23  (.001) | .06 | .25\*  (.006) | <.05 | | -.14  (.001) | .25 | .23  (.007) | .07 | -.17  (.001) | .18 |
| Disruption  Appraisal |  |  |  |  | -.05  (.045) | .66 | -.08  (.006) | .52 |  |  | |  |  | -.06  (.047) | .61 | -.12  (.007) | .34 |
| Confidence  Appraisal |  |  |  |  |  |  |  |  | -.02  (.050) | .85 | | -.13  (.007) | .29 | -.04  (.052) | .76 | -.16  (.007) | .20 |
| R2 | .058 |  | .040 |  | .060 |  | .050 |  | .058 |  | | .054 |  | .061 |  | .069 |  |
| ∆ R2 |  |  |  |  | .002 |  | .010 |  | .000 |  | | .014 |  | .003 |  | .029 |  |
| **End interval (I5)** |  |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |  |
| Teaching  Experience | .04  (.004) | .70 | .11  (.001) | .32 | .07  (.005) | .58 | .18  (.001) | .13 | .09  (.005) | .46 | | .07  (.001) | .58 | .10  (.005) | .43 | .12  (.001) | .33 |
| Disruption  Appraisal |  |  |  |  | .06  (.035) | .60 | .19  (.007) | .12 |  |  | |  |  | .04  (.037) | .76 | .23  (.007) | .07 |
| Confidence  Appraisal |  |  |  |  |  |  |  |  | -.11  (.039) | .38 | | .10  (.008) | .43 | -.10  (.041) | .44 | .16  (.008) | .22 |
| R2 | .002 |  | .013 |  | .005 |  | .053 |  | .012 |  | | .025 |  | .013 |  | .078 |  |
| ∆ R2 |  |  |  |  | .003 |  | .040 |  | .010 |  | | .012 |  | .011 |  | .065 |  |
|  | *Note*. In Model 1, mean standardized HR and mean slopes were predicted only by teaching experience. In Model 2, solely disruption appraisal was added as a predictor. In Model 3, solely confidence appraisal was added as a predictor. In Model 4, all three predictors were considered in concert.  1 We calculated only Model 1 for the pre-teaching interval because the classroom events had not yet occurred in this interval.  \* *p* < .05. | | | | | | | | | | | | | | | | |

**# Discussion**

**## Summary of key findings**

Our study aimed to investigate how data collected from a wrist-worn fitness tracker could shed light on teachers’ stress responses before, during, and after teaching sessions as an effect of stressors, such as classroom disruptions. We assessed teachers’ HR using a Fitbit® fitness tracker over the course of a five-phase lab study, including a micro-teaching unit with potentially disruptive classroom events. Moreover, we examined whether variance in HR measures could be explained by teachers’ teaching experience and self-reported appraisals (disruption and confidence appraisal) of the classroom events.

As expected, teachers’ HR increased before, peaked during, and progressively decreased after the micro-teaching unit. Second, contrary to our expectations, differences in teachers’ HR could not be systematically explained by teaching experience or subjective appraisals of the disruptions.

**## Findings from mapping teachers’ HR over study phases**

Our first research question concerned the effectiveness and suitability of HR measures assessed by wrist-based fitness trackers for mapping teachers’ HR over the course of the five-phase lab study, including the time before, during, and after the potentially stressful micro-teaching unit. Results supported our hypotheses: First, as expected in Hypothesis 1a, mean standardized HR was significantly higher in the micro-teaching unit than in all other phases with large effect sizes (0.82 ≤ *d* ≤ 4.68). This finding is in line with prior studies showing that teachers’ HR varies depending on their activities and encountered stressors, particularly increasing during phases where teachers are in an exposed position [@sperka1995; @scheuch1997psychophysische; @donker2018; @junker2021]. Second, teachers’ mean standardized HR increased before the micro-teaching unit and subsequently declined (Hypothesis 1b). This finding corresponds with results from prior studies that investigated HR trends in teaching-learning situations, showing that HR changes align with activating events or stress-inducing tasks [@Darnell2019; @chalmers2021]. Moreover, researchers found that wearable sensing devices, like smart wristbands, can effectively capture changes in (students’) HR levels as physiological responses during various activities like lectures, self-tests, presentations, and exams [@francisti2023identification]. Third, results revealed that the standardized and mean non-standardized HR values were comparable (see Fig. 3). We used standardized values for all further analyses to ensure that observed differences in HR between individuals were not solely due to inherent differences in baseline HR levels (but see ##Limitations). Taken together, the findings indicate that wrist-worn fitness trackers are a useful tool to map teachers’ HR before, during, and after teaching.

## **Findings from the influence of teaching experience and subjective appraisal ratings on teachers’ HR**

Regarding our second research question, the regression models only partially supported our expectations. Building on the model of teacher stress [@kyriacou1978, see Fig. 2], we hypothesized that more experienced teachers might have better classroom management skills, and thus experience less stress when dealing with classroom disruptions. Not in line with our hypotheses, we found no effects of teaching experience or subjective appraisal ratings on the teachers’ HR in the investigated intervals. This finding suggests that the more experienced teachers were comparably stressed by the teaching demands as the less experienced teachers. At least in the somewhat artificial teaching situation that we created their teaching experience did not provide sufficient resources to effectively prevent stress. Consistent with this finding, recent research suggests that interventions to reduce stress and burnout in teachers need to address multiple levels (individual, individual-organizational, and organizational; @mcintyre2017towards). Merely enhancing teachers’ skills and coping mechanisms might not effectively diminish their stress levels, unless changes are made to the organizational context of schools, including factors such as excessive workloads, resource limitations, and unsupportive administrative practices [@eddy2019single]. Moreover, teachers’ professional experience is not a guarantee for more professional knowledge and skills, as teachers with more professional experience do not perform better than their colleagues with less professional experience in terms of pedagogical-psychological knowledge, which also includes effective classroom management [@kirschner2016professionswissen]. Developing skills from professional experience requires a deliberate practice of teaching “to choose to improve, to learn through […] experience, and to integrate new knowledge into future performances” [@dunn1999deliberate, p. 647].

While we found no systematic effects of teaching experience or subjective appraisals, we did see some interesting patterns of relations. First, teaching experience was predictive of HR differences in the *interview phase*, in the direction that more experienced teachers showed a higher HR (β = .24) and thus, probably experienced higher levels of stress, during the stimulated recall interview. One explanation for the higher HR of experienced teachers could be that age correlated strongly with teaching experience (*r* = .94). For instance, researchers showed that older (female) teachers did not experience a decrease in their HR during periods of low stress levels, from which they concluded that recovery from stress was insufficient in the older teachers [@ritvanen2006responses]. Another explanation could be provided by @alhija2015teacher, who found that more experienced teachers reported more stress due to student misbehavior compared to less experienced teachers as a result of higher burnout symptoms. In other words, experienced teachers may show elevated HRs due to a habitually increased stress level and lower stress tolerance. However, the positive association between teaching experience and HR was found only in the *interview phase*. Thus, it is also possible that more experienced, older teachers found the interview itself to be more stressful. Since the interview entailed watching one’s own lessons together with an experimenter and answering the experimenter’s questions on the classroom disruptions that had occurred during may indeed be considered a potentially stress-inducing situation. Possibly, the more experienced teachers felt more threatened by the interview situation than less experienced teachers.

Second, two findings did support our hypotheses, at least partially: 1) We found teachers who perceived the events as more disruptive had a higher HR (β = .25), i.e., the phase immediately following the micro-teaching unit, and thus probably the most prone to actually show the cumulative effects of the potentially stressful classroom disruptions, and 2) we found a less steep HR increase in teachers with more teaching experience during the *pre-teaching phase* (β = -.27), i.e. in preparation of the micro-teaching unit. The first finding is consistent with the idea that differences in HR, as an indicator of the physiological stress response, can be linked to cognitive appraisals of stressors. The second finding supports the idea that, even though teaching experience guarantees neither superior expertise nor stress resistance, the habits and routines formed by experienced teachers may at least lead to lower arousal levels (e.g., experienced as feeling less nervous and tense) when they anticipate potentially stressful teaching situations.

Taken together, our findings support @wettstein2021 call for the use of ambulatory assessment methods, particularly in the context of classroom disruptions, for gaining a deeper understanding of teacher stress and its impact on both psychological and physiological variables.

**## Limitations and future directions**

While the laboratory setting of the study allowed for a controlled implementation of stressors in a setting that was comparable for all participants, the setting was not an authentic classroom environment. Most importantly, the setting did not include a shared history of the teacher and his or her students, and thus only a very thin basis for establishing a positive teacher-student relationship, which is a core characteristic of effective classroom management [@ruedi2014; @beaty2010]. Nevertheless, conditions were identical for all participants, meaning that even if the HR was influenced by the artificial setting, this was likely the case for all participants and does not limit our general conclusions. In subsequent studies, it would be insightful to assess teachers’ HR in authentic classroom settings over a longer period (e.g., week or month).

Further limitations concern our assessment of teachers’ HR. While our results demonstrate the usefulness of drawing upon easily available HR data from ubiquitous, low-cost, un-intrusive fitness trackers in order to estimate teacher stress, there also are shortcomings of this type of assessment. First, while fitness trackers typically yield HR data, heart rate variability (HRV) data would be an even more accurate indicator of stress [@wettstein2020ambulatory] but is typically not provided by low-cost fitness trackers, at least in currently available models. Depending on the further development of low-cost fitness trackers, however, future studies might consider assessing HRV instead of HR. Second, our design did not include a phase during which participants’ resting HR could be recorded. Resting HR is generally considered an important baseline to determine inter- and intrapersonal differences in cardiovascular health and reactivity [@nanchen2018; @heneghan2019]. However, recording a valid baseline HR requires a resting phase without physical movement or emotional stress, ideally fifteen minutes before the beginning of the activity, which is difficult to impossible to achieve in practice [@sammito2015guideline], e.g., when assessing teacher HR before and during teaching. Thus, our study explored the possibility of substituting baseline HR measurement via z-standardization across participants. As a result, the absolute standardized values of each participant must always be interpreted in the context of the standardization sample, and thus are less interpretable than individual BPM values together with a baseline HR. However, for statistical analyses based on the whole sample, the standardization fulfills the aim of controlling for differences in individual mean HR due to, for example, age-related differences. Finally, depending on the brand and model of fitness trackers used, the measurement of HR may be more or less precise, possibly due to systematic measurement errors. Our study used the same Fitbit® tracker on all participants, but could not compare its results to those of other devices. Research on the reliability of Fitbit® devices for the measurement of HR has proven that this brand is generally accurate in controlled settings and for moderate activity levels [@wallen2016accuracy; @hajj2023; @fuller2020; @jo2016] such as in our study. For example, the Fitbit® fitness tracker showed good measurement accuracy during resting phases [@jo2016; @muggeridge2021measurement] and for activities such as walking, jogging, and running [@hajj2023]. However, some studies indicated that the Fitbit® tracker sometimes underestimates HR at higher exercise intensities such as cycling [@thomson2019heart; @montoye2017comparative; @jo2016; @jachymek2021]. While @chevance2022accuracy concluded in their systematic review that Fitbit’s® the underestimation of HR has an acceptable range. @gagnon2022 stressed that Fitbit® trackers cannot replace an ECG machine when precision is paramount. Despite these considerations, the Fitbit® model was suitable for our study purposes, as the participants did not have to perform any intense activities.

Moreover, our study design allowed us to control the stressfulness of the situation that the teachers experienced, in particular by confronting them with the (scripted) classroom disruptions. However, while the presence of potential stressors is necessary for the emergence of a physiological stress response, many other factors influence whether a stress response occurs, and how it is subjectively registered and expressed by the teacher. We aimed to address at least some of the factors influencing stress responses according to the model of teacher stress [@van2006stress]. However, we missed other factors that potentially also have a substantial effect. For example, we did not assess any personal traits of the participants, such as emotional stability, work commitment, and life satisfaction, which are considered to be potential sources or protective factors of teacher stress [@wettstein2021]. Furthermore, we did not gather any information on the health status of the participants. However, factors such as alcohol consumption, fitness level, cardiovascular diseases, etc. could have influenced physiological responses such as HR and should be taken into account [@sammito2015guideline]. Future research should incorporate additional information collected to account for the fact that human HR is, in addition to the autonomic nervous system and genetic factors, influenced by numerous external factors such as social, personal, psychological, environmental, and behavioral factors [@wang2022].

Furthermore, an important consideration must be made about the primary and secondary appraisal processes, which were measured using single-item scales. It is possible that the ratings were not sensitive enough to assess teachers’ appraisal processes, possibly due to the dynamic interaction between individuals and their environment regarding stress. Stress is not a fixed construct, but a constantly changing affective response, making it difficult to determine process markers for appraisal. Thus, the search for a single satisfactory measure is constrained by the inherent complexity of stress [@lazarus1990theory].

**## Practical implications for teachers and researchers**

Despite the limitations of our study, results suggest that wrist-worn, low-cost, and nonintrusive fitness trackers are a promising tool for recording HR as an indicator of stress in educational and academic settings with practical implications. First, implications for teachers and teaching training will be presented, followed by some central guidelines for researchers to consider when working with fitness trackers.

The increasing availability of HR data from wearable fitness trackers offers teachers the opportunity to self-monitor important mental health indicators such as HR, beyond traditional self-reporting or expensive, intrusive ECG devices. Using fitness trackers could enable teachers to strengthen their self-awareness in stressful situations and allow for early self-intervention such as incorporating mindfulness techniques (e.g., deep breathing or body scans) into daily routines [agyapong2023interventions]. Furthermore, the use of fitness trackers in education could help teachers create a greater awareness of the interplay between teaching practice and physiological and psychological variables. For example, researchers were able to show that an increased HR during teaching was linked to less effective and sometimes confusing prosody patterns such as intonation, pace, and pausing [@tobin2016expression]. Research on mental health has shown that to achieve a regular and meaningful use of fitness trackers as mental health monitoring, it would be essential for participants to understand the purpose of using the fitness trackers as well as the meaning of the data [@ng2018]. Accordingly, to secure a lasting benefit of the technology and use the full potential of fitness trackers for the personal health management of teachers, workshops should be offered, for example, in which teachers are taught the exact use of fitness trackers (how to put on the watch correctly) and how to handle the data (how to download the data, is the handling compliant with data protection, how to interpret HR data, etc.).

Future research could use low-cost and non-invasive devices to accompany teachers in their everyday school practice to gain insight into teachers’ stress experience during teaching. Even in teacher training, wearable fitness trackers could provide new insights into the stress experience of student teachers during internships. The SRI method has proven very insightful for adding the teacher’s own perspective to the recordings of their teaching and physiological data. Evaluating data from fitness trackers, possibly together with video recordings of their lessons, together with teachers or teacher students could provide clues as to which types of situations are experienced as particularly stressful, and to discuss and implement possible stress-reducing measures in teacher training. Accordingly, the combination of subjective self-reported data such as interviews or questionnaires and objective measures such as HR would be an important step towards understanding and possibly preventing the development of stress in the teaching profession.

For researchers wishing to use fitness trackers in data collection, a few practical aspects to consider concerning the planning, data collection, and follow-up procedure of studies.

1. When planning studies with fitness trackers, researchers need to decide which model of fitness tracker is to be used, depending on the research question. Therefore, it must be considered whether the study will be conducted in the laboratory, in a medical environment, or under actual real-world conditions. As already mentioned, conventional fitness trackers should not be used if the focus is on measurement accuracy, such as in medical studies, as they cannot replace ECGs [@gagnon2022]. Moreover, researchers should consider that measurement accuracy also depends on the intensity of the movements performed by the participants during data collection. Findings in some studies indicate that Fitbit® fitness trackers for example showed a decrease in accuracy by underestimating the HR, especially at higher exercise intensities such as cycling [@thomson2019heart; @montoye2017comparative; @jo2016; @jachymek2021]. The systematic review by @fuller2020 provides a detailed overview of studies that used consumer-worn activity trackers between 2000 and 2019 regarding their validity and reliability. Another point that is decisive when choosing a fitness tracker model is the price. Between €30 and up to €1.700 for medical wristbands such as the Empatica®, all price ranges are possible, depending on the research aim. It would be advisable to make sure that the fitness tracker measures HRV rather than just HR. Some models (e.g., from Garmin®) even display an individualized stress level to the user based on the HRV. Before the study, it should also be considered that the data collected with fitness trackers is health data, i.e., sensitive data. This means that researchers have to ensure that the data is processed anonymously and no personal information about the participants is provided, at most belonging to a specific sample (e.g., teachers).
2. Before and during data collection, decisions must be made regarding the circumference, attachment, and placement of the fitness tracker. The circumference depends, for example, on the age of the participants. Thus, studies conducted with children should take into account the small wrist size when attaching the band. When putting on the fitness tracker, attention must also be paid to whether it is attached to the dominant or non-dominant wrist, as this can influence the step count. In terms of placement, researchers should note that fitness trackers need to be placed differently. For example, the Fitbit® Charge 4 should be placed a finger’s width above the wrist bone for daily movement and two finger widths above the wrist bone for exercise, as the wrist is bent more frequently, which can influence the HR signal. In addition, researchers should of course ensure that the fitness tracker is regularly disinfected, especially when used with different participants. It is also important to check that the battery is fully charged each time, that the latest version is loaded on the app and the fitness tracker has been synchronized with the app before recording data to avoid unnecessary loss of data. Moreover, if researchers want to investigate parameters in different time intervals as in our study (e.g., HR in lessons vs. breaks during the school day), it would be advisable to synchronize the fitness tracker with other watches to be able to determine the on- and offset of certain intervals/ time of interests.
3. As far as the further procedure for processing the data is concerned, researchers should ensure that the raw data of the physiological measurements are available for further analysis. For the Fitbit® HR measurements, for example, the raw data can be downloaded from a URL in the form of .csv files. However, these must be downloaded as soon as possible, otherwise the data will also be lost. During follow-up, it is also important to ensure that the data was reliably collected at the intended sampling rate. The model we used states that the fitness tracker records the heart rate every 1-5 seconds (depending on the movement). In our case, however, the measurement exceeded this sampling rate, so we sometimes only had HR measurements every 15 seconds.

For an additional overview, see @nelson2020guidelines. This paper reviews the use of consumer wearables for cardiovascular psychophysiological measurement, emphasizing the need for standardized reporting and outlining guidelines to address inconsistencies in study design, data processing, and demographic considerations to enhance replicability and accuracy.

**## Conclusion**

This study investigated whether HR data collected from teacher-worn fitness trackers are suitable to explore links with factors such as subjective stressor appraisal, or effects of teaching experience to achieve a more profound comprehension of stressful transactional processes occurring in the classroom. Our results suggest that the widespread availability of HR data via fitness trackers presents opportunities for teachers to self-monitor stress levels for early intervention. Integrating fitness trackers into teacher training and everyday practice could offer valuable insights into stress experiences, facilitating the development of targeted interventions to support educator well-being. In summary, our study contributes to the understanding of stress in educational settings and underscores the potential of wearable technology in advancing research on teacher well-being.

**APPENDIX**

**Figure XX**

*Setting of the 15-minute micro teaching unit. Note. The setting included three actors as the class (left) and a teacher (right).*

Ein Bild, das Mobiliar, Stuhl, Kleidung, Schuhwerk enthält.

Automatisch generierte Beschreibung

**Figure XX**

*Setting of the interview. Note. The experimenter and participant watched the previously taught unit on video.*

Ein Bild, das Mobiliar, Zeichnung, Entwurf, Tisch enthält.

Automatisch generierte Beschreibung

**Figure XX**











1. The fluctuations in the number of seconds in which the HR was measured are due to the participants' movements, meaning that the device could not measure the HR every second. [↑](#footnote-ref-1)
2. Note that the study exceeded the planned duration of two hours for a few participants. To avoid distortions when mapping the HR over the course of the study (see Fig. 3), the endpoint was set at two hours for all participants, even though data from later time points was used in the *end interval* for a few participants. [↑](#footnote-ref-2)
3. Although this procedure does not account for nonmonotonic progressions in individual HR, a graphical evaluation revealed that the linear estimates corresponded well to the majority of the cases (see XX in the supplementary material). [↑](#footnote-ref-3)
4. We used the mean standardized HR as we wanted to explain the mean HR of the intervals and not the HR at the beginning of the interval (x = 0). [↑](#footnote-ref-4)