**# Introduction**

In the educational context, there is a particular interest in finding adequate indicators and quantitatively measurable parameters for teacher stress and burnout [@fisher2011; @ junker2021]. Previous research on teacher stress often focused on the psychological experience of stress using self-report questionnaires with single-item measures [@chaplain2008; @goker2012] or questionnaires with multiple scales [@fimian1990; @liu2020]. However, self-reported data in the measurement of stress raises concerns about the validity and accuracy of causal inferences due to response biases such as social desirability [@razavi2001self] or recall bias [@van2016accuracy]. As self-reported data fail to capture actual physiological stress responses in real-life situations, ambulatory assessment methods using a variety of assessments are recommended, for example, collecting additional physiological measures such as HR as objective data [@trull2013ambulatory; @ wettstein2020ambulatory]. Furthermore, HR as a physiological measurement provides researchers with objective insights into teachers’ affectivity and stress levels without interrupting the teaching process [@donker2018; @runge2020].

To date, the studies measuring teachers’ HR as an indicator of stress in teaching-learning settings mostly use cost-intensive and intrusive electrocardiographs [@sperka1995; @scheuch1997psychophysische; @donker2018; @junker2021; @huang2022class], revealing that teacher-centered activities and typical stressors are leading to an increase in HR. However, the recording of HR in educational contexts would benefit from using inexpensive and non-invasive instruments. Wrist-worn fitness trackers have the potential to be such a tool [@ferguson2015]. In contrast to occasional clinical observations, they allow the collection of fine-grained data over a longer period and are less intrusive than complex medical devices (e.g., electrocardiograms) that require body contact [@godfrey2018z]. Furthermore, in the last decade, commercial wearables such as fitness trackers have become increasingly popular and accepted among the wider population [@park2020user]. In addition to the ease of use, perceived usefulness, and enjoyment [@peng2022acceptance], these devices are equipped with biosensors providing users with physiological (e.g., HR, skin temperature) and behavioral data (e.g., step count, distance walked). The general public is thus offered the opportunity to use low-cost, lightweight devices to monitor their physical activity and health routines in their everyday life.

It can be assumed that teachers also wear personal, private fitness trackers, generating recordings of physiological data that could be a very interesting resource for research on teacher stress. While the use of wearable technology in the field has been explored in various domains, including medicine [@hughes2023wearable; @yetisen2018wearables], sports [@secckin2023review; @ adesida2019exploring], and entertainment [@helmer2009smart; @cciccek2015wearable], research is sparse in educational contexts [@de2017towards]. Although some studies investigate how wearables can help teachers monitor student activity [@quintana2016keeping], there is a notable research gap regarding teachers’ use of wrist-worn wearables.

Given the high stress levels in the teaching profession [@johnson2005experience], fitness trackers could be a valuable tool for analyzing HR and the factors contributing to stress. One of the reasons for teachers´ augmented stress is that they are confronted with a multitude of demands in their everyday work, some of which exceed their resources and therefore make it difficult to cope with immediate stressors such as classroom disruptions [@montgomery2005meta]. However, the extent of the strain depends on the subjective appraisal of a stressor, which involves considerations about available resources to deal with it[@kyriacou2001]. It is, therefore, particularly important for teachers to have sufficient personal and professional resources at their disposal [@cramer2018belastung]. Classroom disruption, for example, are one of the major stressors in teachers´ daily work (XXX), and professional knowledge about effective classroom management reduces the risk of teacher stress [@klusmann2012berufliche]. Teachers’ characteristics such as professional experience in turn have an impact on the development of classroom management skills and thus on the appraisal processes, as these skills develop during professional experience [@ophardt2017klassenmanagement; @wolff2015keeping].

To better understand how stressors like classroom disruptions affect teachers and their stress responses , data from wrist-worn fitness trackers could be used to monitore physiological parameters before, during, and after teaching sessions (cf. @wettstein2021). . This study explored the use of wrist-based fitness trackers as a tool to monitor teachers’ stress during different phases of a micro-teaching unit during which teachers had to deal with classroom disruptions. Physiological data were triangulated with teachers´ appraisal of classroom disruptions, and their teaching experience. The physiological indicator employed in this study was teachers´ heart rate (HR), which can be readily recorded by any fitness-tracker.

**## Fitness Trackers as a Method to Assess HR as an Indicator of Stress**

Wearables (also referred to as wearable devices, wearable computers, or wearable electronics, @cciccek2015wearable) are defined as electronic devices that are either directly worn on the body or loosely attached to a person and integrated into clothing or accessories to serve as a convenient all-in-one solution [@godfrey2018z]. Essentially, wearables such as fitness trackers are designed to be worn continuously by users, gathering data such as location, movements, and vital signs via wireless sensors enabling users to interact with these devices anytime and anywhere [@cheng2017underlying]. These gadgets are characterized by attributes such as hands-free operatable, portable, useful, reliable, practical, multi-functional, mobile, socially acceptable, etc. [@cciccek2015wearable, p. 46].

Recently, fitness trackers have become widely popular and accepted as a mass product by the population [@park2020user]. Several factors contribute to widespread acceptance: Fitness trackers monitor various aspects of physical activity like HR, distance, steps, and calories burned, providing valuable insights into users’ daily activity and cardiovascular health, supporting them in setting personalized fitness and health goals [@nuss2021effects] or providing information about stress levels [@hao2018chrv]. Further advantages are the portable, non-invasive nature of these devices, the ease of use, and especially, the low costs compared to complex laboratory methods for determining vital parameters such as HR, blood pressure, skin temperature, or physical activity [@hajj2023].

The blend of these factors makes fitness trackers an ideal technology for use not only in healthcare, entertainment, and fitness [sinha2019taxonomy] but also in education as they offer added benefits for formal and informal learning environments for both students and teachers [@de2017towards]. However, in most studies, the focus is on students, especially to detect students’ skills and enhance their performance [@koutromanos2020use]. Despite the enormous potential of wearables, few studies investigate recorded parameters and in particular their significance for teachers.

One important health parameter assessed by nearly all wrist wearables is HR [@scalise2018wearables]. HR indicates the number of heartbeats within one minute, and is typically expressed as beats per minute (BPM) [@hottenrott2007]. HR can be detected and measured using sensors based on electrocardiogram (ECG) or phonocardiogram (PCG) [@mukhopadhyay2017wearable]. Another uncomplicated and inexpensive technique to measure HR via fitness trackers is photoplethysmography (PPG) [@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].Therefore, an increase in HR can be regarded as an indicator of increasing stress, and a decrease as an indicator of decreasing stress [@kyriacou1978].

**## Teacher Stress and Important Resources**

The teaching profession is one of the most stressful professions compared to other occupational groups, facing a variety of stressors during everyday work [@smith2000; @herman2020; @schult2014belastet]. According to @kyriacou1978, teacher stress can be defined as

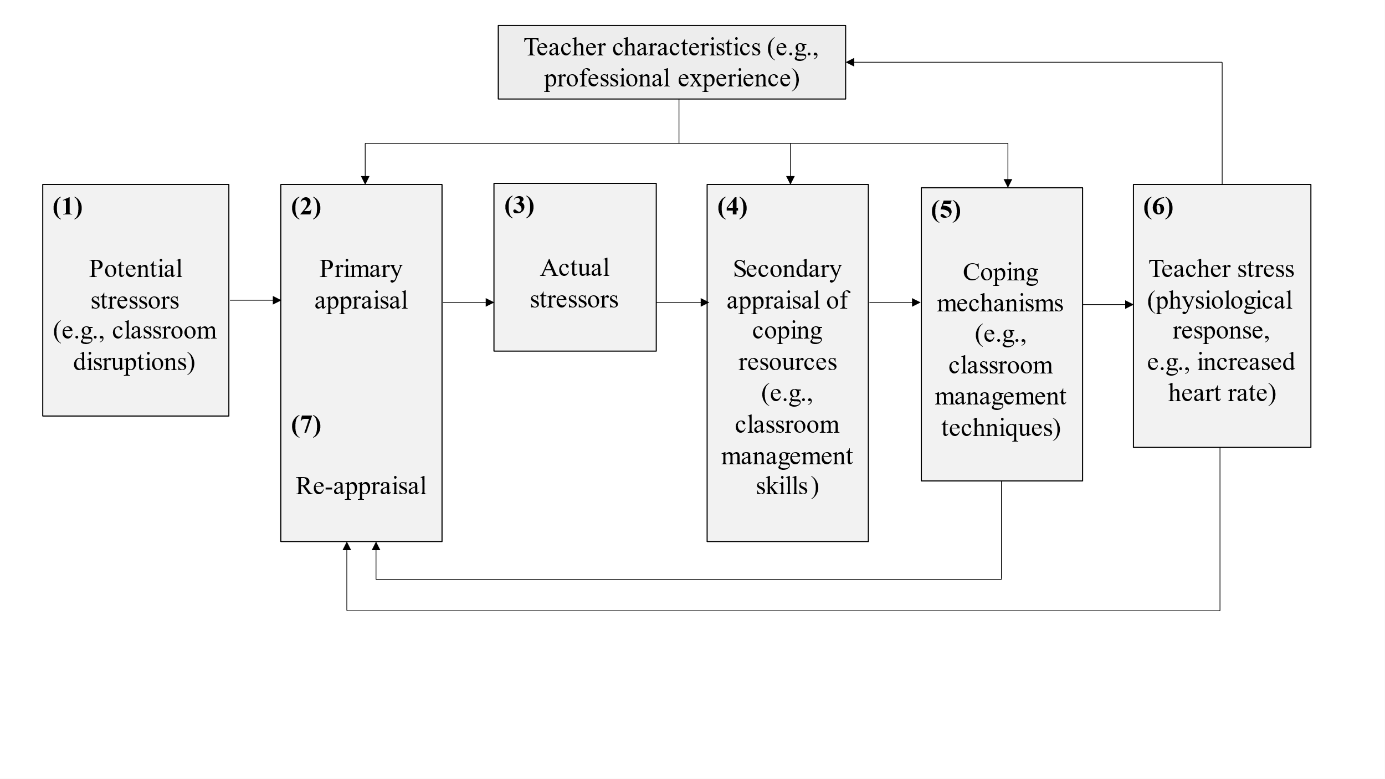
“a response of negative affect […] usually accompanied by potentially pathogenic physiological and biochemical changes (such as increased heart rate […]) resulting from aspects of the teacher’s job and mediated by the perception that the demands made upon the teacher constitute a threat to his self-esteem or well-being and by coping mechanisms activated to reduce the perceived threat.” (p. 2).

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 [@dick2013belastung].

In general, the transactional stress model [@lazarus1990theory] highlights the interaction between an individual and the environment, whereby stress refers to any event that exceeds a person’s adaptive resources. It has been shown that there are important connections between stressors and resources experiences by teachers on the one hand, and stress-induced health issues on the other hand [@krause2013messung]. As we are interested in how specific classroom events affect teacher stress, we adapted the transactional stress model to that type of situation, based on a representation of the model first proposed by [van2006stress]; this working model is depicted in Fig. 1.

**Figure 1**

A Model of Teacher Stress (Adapted From van Dick 2006, p.37, Modified by the Author)



A When potential stressors (e.g. classroom disruptions) occur during teaching (1), teachers subjectively appraise how disruptive the event is (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 stressor (4). In the best case, teachers have both external (e.g., supportive colleagues) and internal (e.g., classroom management skills) resources at their disposal. A lack of resources and coping mechanisms can lead to negative personal and vocational consequences such as burnout, high turnover, and premature retirement [@jalongo2006; @unterbrink2007; @aloe2014]. These correlations are not surprising, as teaching is characterized as multidimensional, simultaneous, immediate, unpredictable, public, and shared [@doyle2013ecological]. Lessons and, in particular, classroom disruptions are unpredictable and multifaceted. All of these circumstances place several demands on teachers and require a high level of knowledge and competence from the teacher [@klieme2008concept]. These professional competencies encompass, among other things, teachers’ specific knowledge and skills about classroom management and can be understood as a diverse toolbox of strategies, techniques, and measures for the teacher to navigate the challenging environment of the classroom [@konig2016teacher].

Additionally, during both primary and secondary appraisal processes, teachers’ characteristics, such as teaching experience (see Fig. 1, upper box), play a particularly important role and have a decisive influence on classroom management skills. Particularly teachers with less teaching experience are overwhelmed by the simultaneity and complexity of teaching [@ophardt2017klassenmanagement; @wolff2015keeping; @ klusmann2012berufliche]. Skills in dealing with teaching events are closely related to the cognitive load of teachers. These classroom management skills develop, among other things, through growing teaching experience, as teachers attempt to cluster experienced classroom events into patterns and formulate appropriate action alternatives. According to @wolff2021classroom, such cognitive processes can be understood as mental classroom management scripts. Especially for beginning teachers, the teaching profession seems to be very demanding and stressful [@schmidt2017makes]. During the first five years in the teaching profession, between 40 and 50 percent of beginning teachers change careers for a variety of reasons such as disciplinary problems with students [@ingersoll2003]. @fisher2011 investigated the extent to which age or teaching experience and job dissatisfaction are associated with an increased risk of burnout and stress among teachers. The results revealed that teachers with less professional experience had higher burnout scores and that years of professional experience, burnout, and satisfaction in the teaching profession are statistically significant predictors of teacher stress.

Based on the evaluation of resources and their characteristics, teachers will try to cope with classroom disruptions (5) and, for example, use classroom management strategies to stop the disruption. If teachers are unable to cope, they experience stress (6). Teacher stress is mainly characterized as a reaction to negative affect (e.g., anger or depression), which is generally followed by other symptoms that can be seen as reactions to teacher stress. The response of negative affect can be among other things a physiological stress reaction, such as an increased HR [huang2022class; @kyriacou1978]. Based on the resource appraisal and the successful or unsuccessful coping with the stressor, the stressor will be appraised again (7).

**## HR in Teaching-Learning Contexts**

ECG studies have revealed that HR can be used to map changes in HR during teaching onto stressors experienced by teachers. For example, HR increased during teacher-centered activities when teachers had to take a leading position in the student-teacher interaction [@sperka1995; @scheuch1997psychophysische; @donker2018; @junker2021]. @sperka1995 for example recorded the HR of 16 pre-service teachers during their first lesson. The results showed significantly increased psychophysiological activation in terms of an increased HR during teaching. The activation effect was particularly prominent at the beginning of the lesson and decreased over the course of the lesson. The authors interpret this result as showing how pre-service teachers’ active coping processes, i.e, the active management of the interaction with the students, helped the teachers regulate their HR downwards. Other ECG studies identified typical stressors predicting increased HR values, such as class size [@huang2022class], or low student engagement and motivation [@junker2021]. For example, @junker2021 recorded the HR of 40 teachers using an ambulatory monitoring system during a real classroom lesson. They provided evidence that teacher stress caused by stressors such as low student engagement and motivation, or teacher-centered acitivies, lead to an increase in 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 using wrist-worn devices during lectures. The analysis revealed a constant decrease in HR from the beginning to the end of a lecture, whereas the HR peak was reached during active learning sessions. @chalmers2021 examined the usability of the average HR, measured with a Fitbit fitness tracker, to identify physiological changes during stress-inducing tasks. in a study with a total of 60 participants,. The average HR increased significantly between the resting and stress phases. Even though the participants in these studies were learners, not teachers,, the results are relevant to the study of teacher stress using wearable devices, because the studies showed that a) HR can be effectively recorded using fitness trackers during 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 wrist-worn wearables during teaching: @runge2020 used a Fitbit fitness tracker to assess HR as an indicator of stress in n=4 teachers. They used the fitness trackers’ recordings to differentiated between teachers reporting higher or lower levels of stress. In particular, it was found that the combination of a high number of steps, a high HR, and short sleep was an indicator of stress, and that poor student behavior was the stressor perceived most frequently by the teachers. It should 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 the activity and stressors they experience, whereby teacher-centered phases, in particular, led to an increase in HR [@sperka1995; @scheuch1997psychophysische; @donker2018; @junker2021]. Furthermore, it could be shown that HR as an indicator of stress can be assessed using low-cost, non-intrusive fitness trackers, and that HR increases in activating phases and even before stress occurs [@Darnell2019; @chalmers2021]. However, studies collecting data from teacher-worn fitness-trackers in a large-enough sample to explore links with factors such as subjective stressor appraisal, or effects teaching experience, are still lacking. Our study aims at closing this research gap.

**## Present Study**

The data analyzed for the present study were obtained from teachers and student teachers who participated in a lab study, as part of a larger project targeting the development of classroom management in teachers.

As part of the larger project, participants cam to the lab individually, and each taught a 15-minute, self-prepared micro-teaching unit to a “class” of three actors (trained student assistants). These actors performed nine, possibly 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; also see Fig## for a depiction of the laboratory setting of the micro-teaching unit). The actors received standardized instructions on a screen (only visible to them, not to the participant) to perform a classroom event every one and a half minutes, and they performed the same scripted disruptions for all participants. While teaching, participants wore eye-tracking glasses, and additionally, their lessons were recorded by cameras.

The micro-teaching unit, with its unfamiliar setting and the scripted disruptions of participants’ teaching flow, was potentially stressful. Thus, we were particularly interested in the development of participants HR before, during, and after this micro-teaching unit. Wrecorded HR data in , with a total duration of approximately two hours In the *teaching phase*, participants taught the micro-teaching unit and experienced the classroom disruptions. In the *post-teaching phase*, participants answered several questionnaires in a comfortable seated position. Next, in the *interview phase*, they watched the video of their 15-minute unit, rated the disruptive classroom events, and answered open questions. Finally, in the *end phase*, participant answered another questionnaire. These conditions were identical for all participants. During the entire study, participants wore a fitness tracker on their wrist.

The goals 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 a five-phase lab study, including the time before, during, and after a potentially stressful micro-teaching unit.

First, regarding the overall HR trend during the entire course of the study, we expected participants´ HR to gradually increase during the pre-teaching phase, to peak during the teaching phase, and to gradually decrease during the remaining time interval. In addition, we examined whether z-standardization of the participants’ mean 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 mean HR values.

Second, we selected five corresponding intervals with a length of ten minutes from each of the five phases and examined the levels of and the changes in HR during these intervals. We presumed the highest HR levels in the micro-teaching unit. Regarding HR changes, we expected an increase at the beginning of the study and a decrease in the following phases.

(2) The second research goal was to examine whether variance in HR measures could be explained by participants’ teaching experience (because, presumable, more experienced teachers might have better classroom management strategies, and thus better resources for coping), and/or by their self-reported subjective appraisals of classroom events (specifically, how disruptive it was, and how confident they felt in their coping). We expected all three variables (teaching experience, disruption appraisal, and confidence appraisal) to be significant predictors of HR in the different phases.

**# Method**

**## Participants**

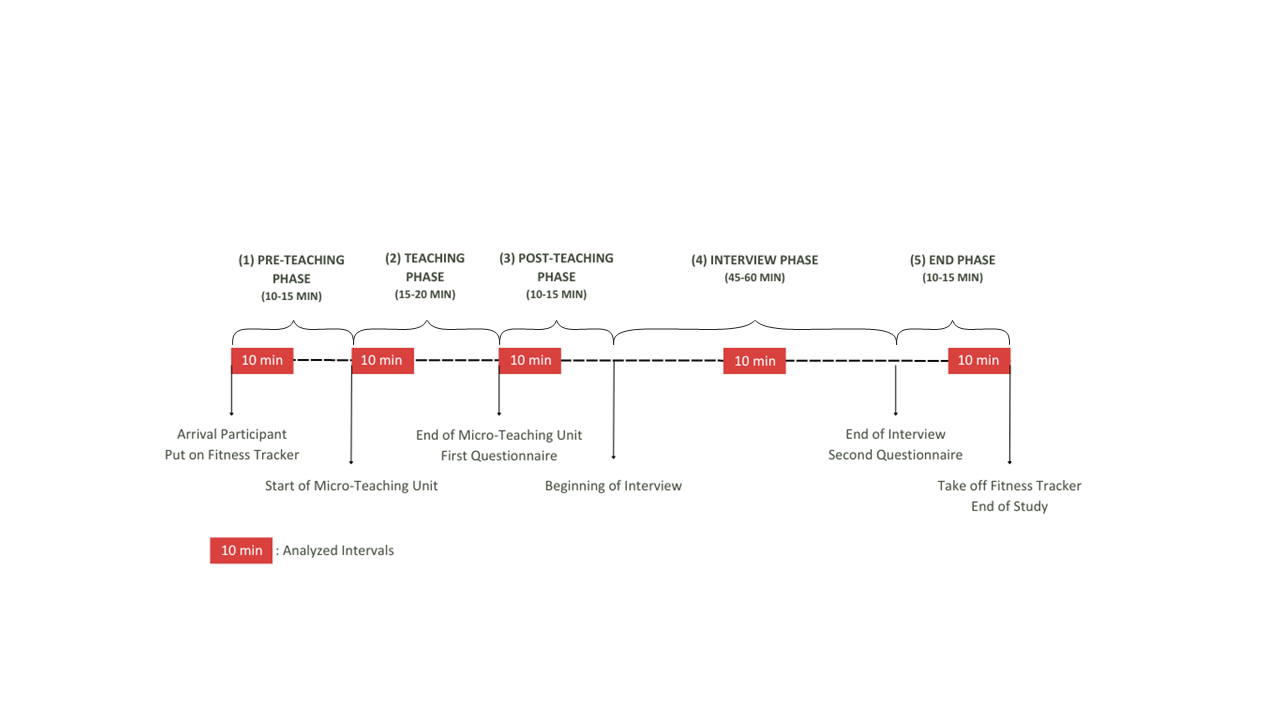
The sample consisted of *N* = 84 pre- and in-service teachers from Germany, who were recruited via personal contact, email lists, and flyers. The data of three participants was lost due to failed data transmission, yielding an analysis sample of *n* = 81 (*n* = 52 women, *n* = 29 men). Participants had 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 and intention of the study before testing. Participation was voluntary and only took place after written consent had been given.

**Figure 2**

*Procedure of the two-hour-long study.*

**

Each participant came to the lab for a period of approximately two hours in total, and each underwent the same phases: *pre-teaching phase*, *teaching phase*, *post-teaching phase*, *interview phase*, and *end phase* (please refer to Fig. 2 for a timeline). 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. The teaching unit was video-recorded and lasted about 15-20 minutes. Participants spent this time mostly standing or slowly walking around. After having completed the micro-teaching unit, in the *post-teaching phase*, participants were seated at a desk and filled in questionnaires. Moreover, in the post-teaching phase, the participants were allowed to take a break after teaching, for example, to use the restroom, drink, or rest. This phase took approximately 10-15 minutes. In the *interview phase*, the participants then watched the video of their own teaching together with the experimenter. While doing so, they were given a Stimulated Recall Interview (SRI), during which they answered questions about their subjective appraisal of the classroom events (see instrument section; also see Fig## in the supplementary material for a depiction of the interview setting). The interview lasted about 45-60 minutes and the participants’ position was seated. The *end phase* lasted about 10-15 minutes and participants answered in a seated position another questionnaire irrelevant to this study.

**### Questionnaire**

In the *post-teaching phase*, the teachers answered questionnaires: 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 that is irrelevant to the present study. All in all, completion of the questionnaires took about 10 minutes.

**### Stimulated Recall Interview**

The SRI took place in the *interview phase*, in which the participants watched their recorded eye-tracking video of the lesson from the ego perspective indicating the participants’ gaze point. The experimenter stopped the video each time one of the nine classroom events happened and asked a total of eight questions, five of which were open and three closed. We assessed – among other questions irrelevant to this study – with two closed questions the teachers’ subjective appraisal of the classroom events that took place during the *teaching phase* in terms of how subjectively disruptive they were (disruption appraisal) and how confident the participants felt dealing with them (confidence appraisal) with one item each. Accordingly, teachers indicated their subjective amount of disruption and confidence for each of the nine classroom events on an 11-point rating scale, ranging from 0 (not at all) to 10 (extremely). The SRI lasted 45-60 minutes on average.

**## Measures**

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

The anonymous HR data was synced via Bluetooth to a commercial Fitbit account. Subsequently, the intraday second-by-second data was exported as a CSV file for each session using the open-source software PulseWatch (Ricci, n.d.), and linked to the participant. To account for individual differences in the baseline HR, we first z-standardized the BPM values from the unstandardized mean HRs. To do this, we first calculated the means and standard deviations of the unstandardized mean HRs based on all values that were available for an individual. Subsequently, we calculated the difference of all measures to the mean and divided these values by the standard deviation. As a consequence, the resulting standardized mean HR values can be interpreted as differences from the overall HR mean in standard deviation units.

Since we aimed to explore teachers’ HR between study phases, we decided to aggregate HR over a typical interval within each phase. To keep intervals comparable in duration, we selected intervals with a length of 10 minutes each. 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 (1) pre-teaching interval comprised the first 10 minutes after the fitness tracker had been put on. The (2) teaching interval 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 (3) post-teaching interval started immediately after the end of the teaching unit. The (4) interview interval was defined as the mid-10 minutes between the end of the teaching unit and the time point where the fitness tracker was taken off. This definition ensured that all participants were being interviewed during this interval. The (5) end interval 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).

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

The subjective disruption and confidence appraisals assessed in the SRI on an 11-point rating scale were averaged across the nine classroom events as we were not interested in individual classroom events, but only in the expected mean level of arousal during the *teaching phase*. Regarding the model (see Fig. 1), the disruption appraisal was used to assess the evaluation of the stressor (see Fig. 1, box 2). The confidence appraisal, in contrast, referred to the resources available for coping with the stressors (see Fig. 1, box 4).

**## Data analysis and Hypotheses**

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

We assumed the highest HR level in the teaching interval (I2) and lower levels in all other intervals(\*\*Hypothesis 1a\*\*). For testing Hypothesis 1a, we initially conducted a one-way ANOVA with repeated measures as an omnibus test. The dependent variable comprised the standardized HR mean for each interval. To identify the highest HR level, we subsequently conducted *t*-tests with planned contrasts as post-hoc tests, accompanied by the effect size *d* [@cohen1988new]. Specifically, we tested the differences between the teaching interval (I2) and the other four intervals.

Note that mean HR was calculated at the subject level of *n* = 81 participants (see Table 1), whereas the mean slope and mean intercept estimates are based on all values at all measurement time points (see Table 2).

Further, we expected an increase of participants HR while they were preparing for teaching during the pre-teaching interval (I1) , and we expected a decrease of participants´ HR during all of the following intervals, because of habituating to (I2) / recovering from (I3-5) the stressful teaching phase(\*\*Hypothesis 1b\*\*). In testing Hypothesis 1b, we conducted a linear estimation of the increase or decrease in HR 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.[[1]](#footnote-3)

Because more experienced teacher potentially have better resources for coping with stressful classroom events, we expected lower HR levels and less steep HR changes for teachers with more teaching experience, in particular during the teaching interval (\*\*Hypothesis 2a\*\*). To test this hypothesis, we investigated the effect of teaching experience on participants’ HR levels and HR changes for each of the five intervals using linear regression models with teaching experience as the sole predictor (Hypothesis 2a).

We expected higher HR levels and steeper HR changes for teachers who felt more disrupted by the enacted classroom events, regardless of their teaching experience (\*\*Hypotheses 2b\*\*). At the same time, we expected lower HR levels and less steep HR changes for teachers who felt more confident in dealing with the events, regardless of teaching experience (\*\*Hypothesis 2c\*\*). To test these hypotheses, we separately augmented the models by either teachers´ disruption appraisal (Hypothesis 2b) or confidence appraisal (Hypothesis 2c), while controlling for shared variance with teaching experience.[[2]](#footnote-4)

Lastly, we hypothesized that each of the three predictors (teaching experience, disruption appraisal, confidence appraisal) uniquely contributes to explaining variance in teachers´ HR levels and changes (\*\*Hypothesis 2d\*\*). To test this hypothesis, we also examined the effects of all three predictors in one linear regression model (Hypothesis 2d).

**# Results**

**## Research goal 1: Mapping HR Over Study Phases**

The first part of our first research goal was to map the participants’ overall HR trend and explore whether z-standardization of participants’ mean 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. displays the unstandardized mean HR in BPM and the standardized mean HR, respectively. Referring to the participants’ overall HR trend, HR initially increased, peaked, and then decreased. Comparing the unstandardized and standardized HR trends revealed a high similarity of the overall courses. For all further analyses, we used participants´ standardized mean HR.

**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 | 90.09/0.041 | 15.76/0.991 | 51/-4.03 | 164/4.56 | |
| (1) Pre-Teaching Interval | 96.28/0.48 | 14.11/0.88 | 56/-3.56 | 139/3.24 | |
| (2) Teaching Interval | 100.80/0.85 | 16.23/0.77 | 63/-2.18 | 164/4.37 | |
| (3) Post-Teaching Interval | 93.61/0.27 | 14.01/0.76 | 60/-2.17 | 150/3.06 | |
| (4) Interview Interval | 82.32/-0.72 | 11.85/0.74 | 51/-2.51 | 132/4.39 | |
| (5) End Interval | 77.95/-1.07 | 11.14/0.57 | 50/-2.68 | 120/2.96 | |
| 1 Please note that *M* and *SD* of the overall course were subject to rounding differences in the statistic software RStudio [@RStudio2020]. | | | | |

**Figure 3**

*Overall Course of the HR with the Unstandardized HR in BPM Shown in Fig. a. and the z-standardized HR Shown in Fig. b.*



*Note:* The shadow around the line represents the 95% confidence interval. The confidence interval shown refers to the HR measurement points during the entire study period.

The second part of our first research goal was to locate the HR peak, testing the hypothesis that HR will peak during the micro-teaching unit (Hypothesis 1a). Repeated measures ANOVA revealed that the standardized HR means of the intervals differed statistically significantly between intervals, *F*(4, 400) = 257.50, *p* < .05, *f* = 1.60 (large effect). Post-hoc contrasts indicated that the standardized mean HR was significantly higher in the teaching interval (I2) compared to the pre-teaching interval (I1), *t*(1) = 32.71, *p* < .05, *d* = 0.82 (large effect). Moreover, the standardized HR mean of the teaching interval (I2) was significantly higher than in the post-teaching interval (I3), *t*(1) = 32.00, *p* < .05, *d* = 1.34 (large effect), the interview interval (I4), *t*(1) = 453.47, *p* < .05, *d* = 3.37 (large effect), and the end interval (I5), *t*(1) = 511.89, *p* < .05, *d* = 4.68 (large effect). Thus, as hypothesized, HR peaked in the teaching interval (see Fig. 4).

**Figure 4**

*Standardized Mean HR for the Five Intervals*



*Note:* The dotted line represents the grand mean. Error bars represent the 95% confidence interval around the mean.

Next, we examined the HR changes within each interval to test whether we would find a positive slope in the pre-teaching and negative slopes in the post-teaching, interview, and end intervals (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 slope means of the (1) pre-teaching interval was significantly positive, indicating a rising HR for this interval. In contrast, the slope means of the (2) teaching interval and (3) post-teaching interval were significantly negative, indicating a decreasing HR. For the last two intervals, the (4) interview interval and (5) end interval, the slope mean was also 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 Different Intervals for all Individuals*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 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 | .01 |
| (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 Mean Slopes of the Standardized Mean HR for Each Interval*



**## Research Goal 2: Prediction of Standardized Mean HR and Slopes With Teaching Experience and Self-Report Data**

Correlations among HR, teaching experience, disruption appraisal, and confidence appraisal are presented separately for the five intervals in Table 3. Correlations between HR and the other constructs were mostly very small and statistically non-significant. Correlations between teaching experience and appraisals were substantial: more experienced teachers had lower disruption appraisals and higher confi. App.

Teaching experience significantly predicted standardized mean HR only in the (4) interview interval (*b* = .012, *p* < .05, Table 4, Interview Interval, Model 1), indicating slightly higher standardized mean HR for teachers with more teaching experience. This finding is not in line with Hypothesis 2a.

Teaching experience significantly predicted the magnitude of participants’ HR increase only in the (1) pre-teaching interval(*b* = -.004, *p* < .05, Table 4, Pre-Teaching Interval, Model 1), indicating less steep HR changes in teachers with more teaching experience.

Adding the disruption appraisal while controlling for the shared variance with teaching experience (testing \*\*Hypothesis 2b\*\*) revealed a significant effect on teachers’ HR changes for teaching experience as a predictor only in the (4) interview interval (*b* = -.001, *p* < .05, Table 4, Interview Interval, Model 2), indicating less steep HR changes in teachers with more teaching experience while controlling for the disruption appraisal.

When adding the confidence appraisal while controlling for the shared variance with teaching experience (testing \*\*Hypothesis 2c\*\*), teaching experience significantly predicted mean HR only in the (4) interview interval *(b* = .013, *p* < .05, Table 4, Interview Interval, Model 3), indicating higher mean HRs for teachers with more teaching experience.

When considering the effects of the three predictors in concert (testing \*\*Hypothesis 2d\*\*), mean HR was significantly predicted by disruption appraisal in the (3) post-teaching interval (*b* = 0.084, *p* < .05, Table 4, Post-Teaching Interval, Model 4), indicating higher mean HR for teachers who reported higher disruption appraisal (controlling for all other factors). Furthermore, HR changes were significantly predicted by disruption appraisal in the (5) end interval (*b* = .015, *p* < .05, Table 3, End Interval, Model 4), indicating steeper HR changes for teachers who reported higher disruption appraisal (controlling for all other factors).

**Table 3**

*Correlations Between Standardized Mean HR (HR) and the Predictor Variables Teaching Experience (TE), Disruption Appraisal (DA), and Confidence Appraisal (CA) for the Five Intervals*

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | 1 | 2 | 3 |
| (1) Pre-teaching Interval |  |  |  |
| 1 HR | − |  |  |
| 2 TE | − .17 | − |  |
| 3 DA | − .01 | − .36\* | − |
| 4 CA | − .10 | .44\* | − .37\* |
| (2) Teaching Interval |  |  |  |
| 1 HR | − |  |  |
| 2 TE | .11 | − |  |
| 3 DA | − .20 | − .36\* | − |
| 4 CA | .06 | .44\* | − .37\* |
| (3) Post-teaching Interval |  |  |  |
| 1 HR | − |  |  |
| 2 TE | − .04 | − |  |
| 3 DA | .24 | − .36\* | − |
| 4 CA | .04 | .44\* | − .37\* |
| (4) Interview Interval |  |  |  |
| 1 HR | − |  |  |
| 2 TE | .24\* | − |  |
| 3 DA | − .13 | − .36\* | − |
| 4 CA | .09 | .44\* | − .37\* |
| (5) End Interval |  |  |  |
| 1 HR | − |  |  |
| 2 TE | .04 | − |  |
| 3 DA | .04 | − .36\* | − |
| 4 CA | − .07 | .44\* | − .37\* |
| *Note.* HR = Standardized Mean Heart Rate, TE = Teaching Experience, DA = Disruption Appraisal, CA = Confidence Appraisal, \* *p* < .05. | | | |

**Table 4**

*Multiple Linear Regression of Standardized Mean Heart Rate and Slopes Predicted by Teaching Experience, Disruption Appraisal, and Confidence Appraisal for the Five Intervals*

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | *Dependent Variable: Standardized Mean HR and Slopes* | | | | | | | | | | | | | | | | | |
|  | Model 1 | | | | | Model 2 | | | | | Model 3 | | | | | Model 4 | | | |
|  | Std. Mean HR | | | Slopes | | Std. Mean HR | | Slopes | | | Std. Mean HR | | Slopes | | | Std. Mean HR | | Slopes | |
|  | b (SE) | *p* | b (SE) | | *p* | b (SE) | *p* | | b (SE) | *p* | b (SE) | *p* | | b (SE) | *p* | b (SE) | *p* | b (SE) | *p* |
| **(1) Pre-Teaching Interval1** |  |  |  | |  |  |  | |  |  |  |  | |  |  |  |  |  |  |
| Constant | 0.524\* (0.057) | <.05 | 0.106 | | <.05 |  |  | |  |  |  |  | |  |  |  |  |  |  |
| Teaching  Experience | -.008 (.005) | .12 | -.004\* | | <.05 |  |  | |  |  |  |  | |  |  |  |  |  |  |
| R2 | .030 |  | .071 | |  |  |  | |  |  |  |  | |  |  |  |  |  |  |
| **(2) Teaching Interval** |  |  |  | |  |  |  | |  |  |  |  | |  |  |  |  |  |  |
| Constant | 0.813\* (0.057) | <.05 | -0.37\*  (0.014) | | <.05 | 1.150\* (0.227) | <.05 | | -0.099  (0.060) | .10 | 0.779\* (0.349) | <.05 | | -0.111  (0.086) | .20 | 1.274\* (0.471) | <.05 | -0.211  (0.116) | .07 |
| Teaching  Experience | .005 (.005) | .34 | -.000  (.001) | | .83 | .002 (.005) | .73 | | .000  (.001) | .89 | .005  (.006) | .42 | | -.001  (.001) | .57 | .003  (.006) | .67 | -.000  (.001) | .78 |
| Disruption  Appraisal |  |  |  | |  | -.062 (.041) | .13 | | .011  (.011) | .29 |  |  | |  |  | -.065  (.042) | .13 | .014  (.011) | .21 |
| Confidence  Appraisal |  |  |  | |  |  |  | |  |  | .004 (.046) | .92 | | .001  (.011) | .38 | -.014 (.047) | .76 | .014  (.016) | .26 |
| R² | .012 |  | .000 | |  | .040 |  | | .015 |  | .012 |  | | .010 |  | .042 |  | .031 |  |
| ∆ R² |  |  |  | |  | .028 |  | | .015 |  | .000 |  | | .010 |  | .030 |  | .031 |  |
| **(3) Post-teaching Interval** |  |  |  | |  |  |  | |  |  |  |  | |  |  |  |  |  |  |
| Constant | 0.272\* (0.056) | <.05 | -0.058  (0.013) | | <.05 | -0.122 (0.222) | .59 | | 0.008  (0.056) | **.89** | 0.069 (0.343) | .84 | | -0.387  (0.081 | .63 | -0.570 (0.457) | .22 | 0.056  (0.108) | .61 |
| Teaching  Experience | -.002 (.005) | .70 | -.0 | |  | .002 (.005) | .76 | | -.001  (.001) | .52 | -.003 (.006) | .55 | | -.000  (.001) | .91 | -.000  (.006) | .91 | -.000  (.001) | .69 |
| Disruption  Appraisal |  |  |  | |  | .073 (.040) | .07 | | -.012  (.010) | .22 |  |  | |  |  | .084\*  (.041) | <.05 | -.013  (.010) | .20 |
| Confidence  Appraisal |  |  |  | |  |  |  | |  |  | .027 (.045) | .55 | | -.003  (.011) | .80 | .051 (.046) | .27 | -.006  (.011) | .60 |
| R2 | .002 |  | .001 | |  | .043 |  | | .020 |  | .006 |  | | .002 |  | .058 |  | .023 |  |
| ∆ R2 |  |  |  | |  | .041 |  | | .019 |  | .004 |  | | .001 |  | .056 |  | .022 |  |
| **(4) Interview Interval** |  |  |  | |  |  |  | |  |  |  |  | |  |  |  |  |  |  |
| Constant | 0.793\* (0.062) | <.05 | -0.014  (0.009) | | .13 | -0.684\* (0.252) | <.05 | | 0.018  (0.038) | .64 | -0.721 (0.382) | .06 | | 0.043  (.054) | .43 | -0.541 (0.522) | .30 | 0.097  (0.073) | .19 |
| Teaching  Experience | .012\* (.006) | <.05 | -.001  (.001) | | .07 | .011 (.006) | .06 | | -.001\*  (.001) | <.05 | .013\* (.006) | .04 | | -.001  (.001) | .24 | .012  (.007) | .07 | -.001  (.001) | .17 |
| Disruption  Appraisal |  |  |  | |  | -.020 (.045) | .66 | | -.006  (.007) | .40 |  |  | |  |  | -.024  (.047) | .61 | -.008  (.007) | .27 |
| Confidence  Appraisal |  |  |  | |  |  |  | |  |  | -.010 (.050) | .85 | | -.008  (.007) | .29 | -.016 (.052) | .76 | -.009  (.007) | .20 |
| R2 | .058 |  | .040 | |  | .060 |  | | .050 |  | .058 |  | | .054 |  | .061 |  | .069 |  |
| ∆ R2 |  |  |  | |  | .002 |  | | .010 |  | .000 |  | | .014 |  | .003 |  | .029 |  |
| **(5) End Interval** |  |  |  | |  |  |  | |  |  |  |  | |  |  |  |  |  |  |
| Constant | -1.076\* (0.049) | <.05 | -0.017  (0.010) | | .07 | -1.176\* (0.199) | <.05 | | -0.089  (0.040) | .03 | -0.811\* (0.300) | <.05 | | -0.075  (0.058) | .20 | -0.897\* (0.411) | <.05 | -0.184\*  (0.077) | <.05 |
| Teaching  Experience | .002 (.004) | .70 | .001  (.001) | | .32 | .003 (.005) | .58 | | .001  (.001) | .12 | .004 (.005) | .46 | | .000  (.001) | .63 | .004 (.005) | .43 | .001  (.001) | .35 |
| Disruption  Appraisal |  |  |  | |  | .019 (.035) | .60 | | .013  (.007) | .07 |  |  | |  |  | .011 (.037) | .76 | .015\*  (.007) | <.05 |
| Confidence  Appraisal |  |  |  | |  |  |  | |  |  | -.035 (.039) | .38 | | .008  (.008) | .32 | -.032 (.041) | .44 | .011  (.008) | .15 |
| R2 | .002 |  | .013 | |  | .005 |  | | .053 |  | .012 |  | | .025 |  | .013 |  | .078 |  |
| ∆ R2 |  |  |  | |  | .003 |  | | .040 |  | .010 |  | | .012 |  | .011 |  | .065 |  |
|  | *Note*. Coefficients are unstandardized with standard errors in parentheses. Effects of teaching experience and appraisals on teachers’ standardized mean HR are displayed for the five intervals.  In Model 1, standardized mean HR was 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. | | | | | | | | | | | | | | | | | | |

**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. 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)
2. the HR levels and HR changes were only predicted in the (2) teaching interval, the (3) post-teaching interval, the (4) interview interval and the (5) end interval with the disruption and confidence appraisal, i.e., not in the (1) pre-teaching interval. [↑](#footnote-ref-4)