New approaches to teachers' experience of stress: Do heart rate measurements with fitness trackers provide an efficient, inexpensive, and robust measurement method?

Mandy Klatt[a,1,*], Peer Keßler[a,2], Gregor Kachel[a,1,1], Christin Lotz[a,1], Anne Deiglmayr[a,3]

[a] *Empirische Schul- und Unterrichtsforschung, Marschnerstr. 29, Leipzig, 4109, Sachsen, Germany*
[b] *Department, Deutscher Pl. 6, Leipzig, 4103, Sachsen, Germany*

Abstract

One or two sentences providing a **basic introduction** to the field, comprehensible to a scientist in any discipline.

Two to three sentences of **more detailed background**, comprehensible to scientists in related disciplines.

One sentence clearly stating the **general problem** being addressed by this particular study.

One sentence summarizing the main result (with the words "**here we show**" or their equivalent).

Two or three sentences explaining what the **main result** reveals in direct comparison to what was thought to be the case previously, or how the main result adds to previous knowledge.

One or two sentences to put the results into a more **general context**.

Two or three sentences to provide a **broader perspective**, readily comprehensible to a scientist in any discipline.

XXX In this proof-of-concept study, we aimed to advance the field of teacher stress by collecting heart rate data with wrist-worn devices and testing a methodology that has the potential to provide more insights on the non-invasive assessment of teacher stress. XXX

Keywords: heart rate, photoplethysmography, wearable electronic device, teaching

---

1. Introduction

In education, there's interest in identifying reliable indicators for teacher stress and burnout (Fisher, 2011; Junker et al., 2021). Previous research often relied on self-report questionnaires (Chaplain, 2008; Liu and Yan, 2020), which may be prone to biases like social desirability (**?**) or recall errors (**?**). To overcome these limitations, ambulatory assessment methods are recommended (**??**), such as measuring physiological parameters like heart rate (HR), offering objective insights into teachers' stress levels without disrupting their teaching (Donker et al., 2018; Runge et al., 2020).

Current studies on teachers' HR in teaching settings often rely on expensive and invasive electrocardiographs (Sperka and Kittler, 1995; Scheuch and Knothe, 1997; Donker et al., 2018; Junker et al., 2021; Huang et al., 2022), showing that teacher-centered activities and common stressors lead to increased HR. However, using affordable and non-invasive instruments like wrist-worn fitness trackers (Ferguson et al., 2015) could enhance HR recording in educational contexts. Unlike clinical devices, fitness trackers could offer continuous

* Corresponding author

Email addresses: `mandy.klatt@uni-leipzig.de` (Mandy Klatt), `Peer@example.com` (Peer Keßler), `gregor@example.com` (Gregor Kachel), `christin@example.com` (Christin Lotz), `anne@example.com` (Anne Deiglmayr)

[1] ConceptualizationWriting - Original Draft PreparationWriting - Review & Editing

[2] Writing - Original Draft PreparationWriting - Review & Editing

[3] Supervision

and less intrusive data collection over time (Godfrey et al., 2018), aligning with the increasing popularity and acceptance of wearables among the general population (Peng et al., 2022). These devices, equipped with biosensors, provide users with physiological and behavioral data, offering an accessible means for monitoring physical activity and health in daily 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 [Hughes et al. (2023); @ adesida2019exploring; helmer2009smart], research is sparse in educational contexts (de Arriba-Pérez et al., 2017). Although some studies investigated how wearables can help teachers monitor student activity (Quintana et al., 2016), there is a notable research gap regarding teachers' use of wrist-worn wearables. Given the high stress levels in the teaching profession (**?**), 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 (**?**). However, the extent of the strain depends on the subjective appraisal of a stressor, which involves considerations about available resources to deal with it (Kyriacou, 2001). It is, therefore, particularly important for teachers to have sufficient personal and professional resources at their disposal (**?**). Classroom disruptions, for example, are one of the major stressors in teachers' daily work (**?**Aloe et al., 2014) and professional knowledge about effective classroom management reduces the risk of teacher stress (**?**). 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 (**??**).

To better understand how stressors like classroom disruptions affect teachers and their stress responses, data from wrist-worn fitness trackers could be used to monitor physiological parameters before, during, and after teaching sessions (Wettstein et al., 2021). 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' HR, which can be readily recorded by any fitness tracker. Teachers' use of wrist-worn fitness trackers in educational research holds transformative potential as obtained fine-grained data and underscores the critical need to monitor teachers' health as they navigate the stressful landscape of the classroom.

1.1. Fitness Trackers as a Method to Assess HR as an Indicator of Stress

Wearables, defined as electronic devices worn directly on the body or integrated into clothing or accessories, serve as versatile solutions (Godfrey et al., 2018), gathering data like location, movements, and vital signs (Cheng and Mitomo, 2017). Fitness trackers, a popular wearable type (Park, 2020), offer insights into physical activity and cardiovascular health, supporting personalized fitness goals (Nuss et al., 2021) and stress management (Hao et al., 2018). Their affordability and ease of use make them ideal not only for various domains including 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 (de Arriba-Pérez et al., 2017). Yet, few studies focus on their significance for teachers.

One important health parameter assessed by nearly all wrist wearables is HR (**?**). HR indicates the number of heartbeats within one minute and is typically expressed as beats per minute (BPM) (Hottenrot, 2007). At rest, the average HR of adults typically ranges from 60 to 80 BPM (Sammito et al., 2015). HR can be detected and measured using sensors based on electrocardiogram (ECG) or phonocardiogram (PCG) (Mukhopadhyay and Islam, 2017). Another uncomplicated and inexpensive technique to measure HR via fitness trackers is photoplethysmography (PPG) (**?**). This optical method assesses HR by flashing green or red lights to measure changes in blood volume (Allen, 2007).
Physiologically, HR is regulated and influenced on short-time intervals by the sympathetic and the parasympathetic nervous system (Pham et al., 2021). An increase in the activity of the sympathetic system results in HR being speeded up ("fight or flight") (Taelman et al., 2009). In contrast, increased activity of the parasympathetic has the effect of slowing down the HR ("rest and digest") (Battipaglia and Lanza, 2015).

Therefore, an increase in HR can be regarded as an indicator of increasing stress, and a decrease as an indicator of decreasing stress (Kyriacou and Sutcliffe, 1978).

1.2. 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 (Smith, 2000; Herman et al., 2020; **?**). According to Kyriacou and Sutcliffe (1978), teacher stress can be defined as a negative affective response, often 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 (Lazarus and Launier, 1981; Lazarus and Folkman, 1984), which was modified and tailored to the teaching-learning environment by Kyriacou and Sutcliffe (1978).

In general, the transactional stress model (**?**) 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 experienced by teachers on the one hand, and stress-induced health issues on the other hand (**?**). 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.

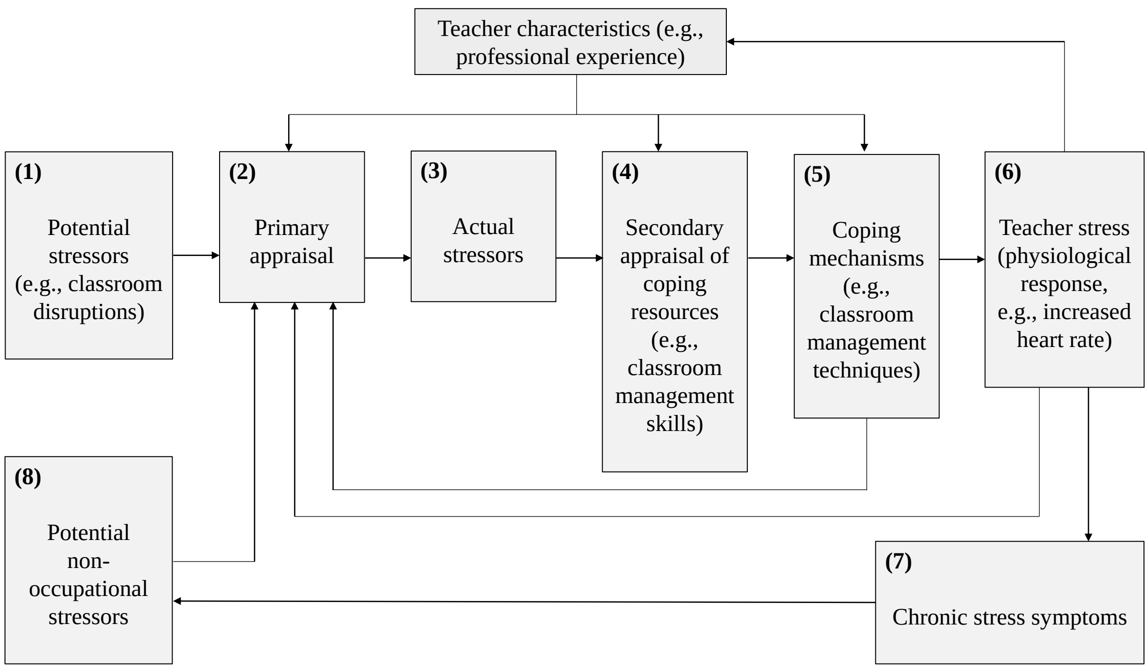


Figure 1: A model of teacher stress (adapted from van Dick 2006, p.37, modified by the author)

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). Based on this evaluation of resources and their characteristics such as professional experience, teachers try to manage classroom disruptions, by utilizing classroom management strategies (5). In cases where coping fails, stress ensues, often accompanied by physiological reactions like increased HR (6). Following resource appraisal and coping outcomes, the stressor is re-evaluated (7).

The appraisal processes of resources play a crucial role as teachers' assessments of resources and coping mechanisms, both internal (e.g., classroom management skills) and external (e.g., supportive colleagues), significantly impact their stress levels. When resources are lacking and coping fails, negative consequences like burnout and high turnover can arise (Jalongo and Heider, 2006; Unterbrink et al., 2007; **?**). This highlights the importance of teachers' appraisal of their professional competencies, including their ability to manage classroom disruptions effectively (**??**).

As shown in Fig. 1, both primary and secondary appraisals are influenced by teachers' characteristics, e.g., their professional experience, shaping their classroom management skills. As experience grows, teachers develop cognitive scripts for managing classroom events resulting in more complex classroom management skills (**?**). Especially beginning teachers face considerable stress and often feel overwhelmed by the demands of teaching [**?**; **?**; @ klusmann2012berufliche], with many leaving the profession within the first five years (Ingersoll and Smith, 2003). Research shows less experienced teachers are more susceptible to burnout, underscoring the importance of experience and job satisfaction in predicting teacher stress (Fisher, 2011).

1.3. 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 (Sperka and Kittler, 1995; Scheuch and Knothe, 1997; Donker et al., 2018; Junker et al., 2021). Sperka and Kittler (1995) 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 (Huang et al., 2022), or low student engagement and motivation (Junker et al., 2021). For example, Junker et al. (2021) 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, i.e., students displaying a lack of motivation or limited interest in their tasks or classroom activities, or teacher-centered activities, i.e., classroom activities primarily focused on the teacher's actions., leads 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 (**?**Chalmers et al., 2021). **?** 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. Chalmers et al. (2021) 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 a wrist-worn fitness tracker during teaching : Runge et al. (2020) used a Fitbit fitness tracker to assess HR as an indicator of stress in N=4 teachers. They used the fitness trackers' recordings to create a profile for each teacher (to differentiate 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. 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 (Sperka and Kittler, 1995; Scheuch and Knothe, 1997; Donker et al., 2018; Junker et al., 2021). 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 (**?**Chalmers et al., 2021). 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 of teaching experience, are still lacking. Our study aims to close this research gap.

1.4. 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 came 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. 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. We recorded HR data in five phases, with a total duration of approximately two hours: 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 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*, participants 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, we expected the participants to show an initial increase in their HR, followed by a peak during the teaching phase and a decrease for the remaining phases. 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, five representative 10-minute intervals were selected from the five phases (see section ##Measures for a more detailed description of the intervals), which served as the basis for the data analysis for our hypothesis: pre-teaching interval (I_1), teaching interval (I_2), post-teaching interval (I_3), interview interval (I_4), end interval (I_5). We examined the levels of and the changes in HR during these intervals. We assumed the highest HR level in the teaching interval (I_2) and lower levels in all other intervals (**Hypothesis 1a**). Further, we expected an increase in participants' HR while they were preparing for teaching during the pre-teaching interval (I_1), and we expected a decrease in participants' HR during all of the following intervals, because of habituating to (I_2) and recovering from (I_3-I_5) the stressful teaching phase (**Hypothesis 1b**).

(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 lower HR levels for teachers with more teaching experience, particularly during the teaching interval (**Hypothesis 2a**). We expected higher HR levels 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 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 examined the same as for the HR levels again for the changes in HR.

2. Method

2.1. 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 = `N_postwrangling` (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).

2.2. 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. Participants were not rewarded in any way for participating in the 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 (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). In preparation for the 15-minute lesson, the topic and class level could be freely chosen by the teachers. The type of course was to be an introductory lesson and the desired social form required individual work or frontal teaching, without longer video sequences and movement of the actors. The teaching unit 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*, partici-

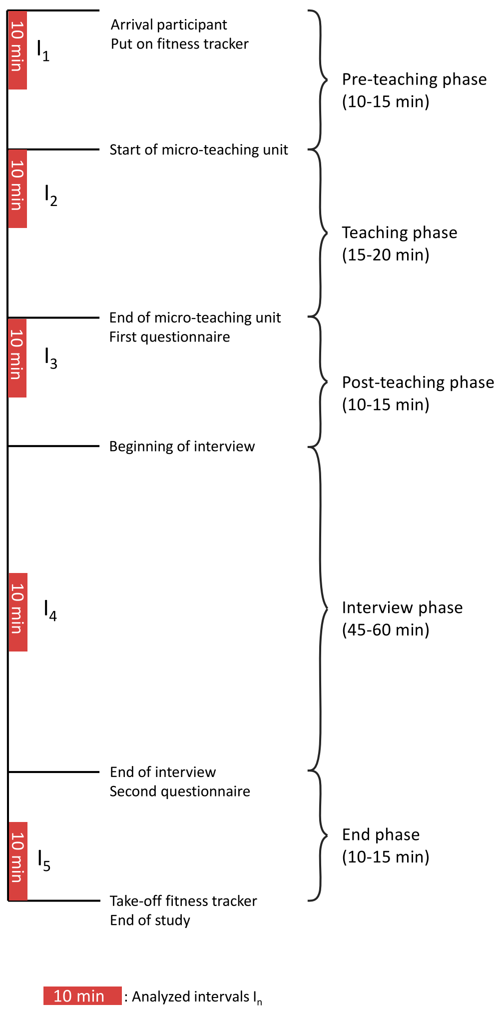


Figure 2: Procedure of the two-hour-long study consisting of five phases with five representative 10-minute intervals

pants were seated at a desk and 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 that is irrelevant to the present study. In the *interview phase*, the participants watched the video of their own teaching together with the experimenter. While doing so, they were given a Stimulated Recall Interview (SRI), 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. 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.

2.3. Measures

2.3.1. Heart Rate Data and Heart Rate Intervals

To measure the teachers' HR, we used a wrist-based fitness tracker. The model was a Fitbit Charge 4. In line with the manufacturer's instructions (, n.d.), the device was attached a finger's width above the participants' nondominant hand's wrist bone. The tracker works by flashing green LEDs hundreds of times per second, using light-sensitive photodiodes to catch the reflected light, and from that information calculating volume changes in the capillaries. From this, the tracker calculates how many times the heart beats per minute. HR measurements are generated at least every 15 seconds[4]. The raw data that can be extracted from the tracker lists the time stamps of all measurements and the estimated HR in BPM for each time stamp.

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.

Since we aimed to explore teachers' HR between different strain phases, we aggregated HR over a 10-minute interval within each phase. Previous research has indicated that 10-minute intervals are a useful duration for analyzing PPG data (**?**). The intervals were selected based on the following rules: The pre-teaching interval (I_1) comprised the first 10 minutes after the fitness tracker had been put on. The teaching interval (I_2) 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 (**??**). The post-teaching interval (I_3) started immediately after the end of the teaching unit. The interview interval (I_4) 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 (I_5) comprised the last 10 minutes before the fitness tracker was taken off.

2.3.2. 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).

[4]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.

2.3.3. 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).

2.4. Data analysis

We conducted all analyses with R (RStudio Team, 2020). Graphics were created using ggplot2 (v3.3.3; Wickham, 2016).

Research goal 1. The first research goal included mapping teachers' HR before, during, and following a micro-teaching unit in the course of a five-phase lab study.

Regarding the teachers' HR trend, we displayed the HR trend over the course of the entire study. For z-standardization as a method to account for individual differences in the baseline HR, we visually compared unstandardized and standardized mean HR trends for the entire two-hours study[ˆ3]. For all further analyses, we used standardized rather than unstandardized HR values. [ˆ3]: 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.

For testing Hypothesis 1a, which examined the HR levels in the different intervals, 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 (**?**). Specifically, we tested the differences between the teaching interval (I_2) 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). For testing Hypothesis 1b, which examined the HR changes within each interval, 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 (**?**) for each interval to estimate intercepts and linear slopes for all individuals which were then averaged across individuals[5].

Research goal 2. In addressing our second research goal, we examined the effects of teaching experience and subjective appraisal of disruptive classroom events on teachers' HR levels during the five phases.

To test hypothesis 2a, we examined the effect of teaching experience on participants' HR levels 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 models by either teachers' disruption appraisal (Hypothesis 2b) or confidence appraisal (Hypothesis 2c), while controlling for shared variance with teaching experience. HR levels were only predicted with the disruption and confidence appraisal after the teaching had taken place, i.e., for all the intervals following the pre-teaching interval (I1). To test hypothesis 2d, we examined the effects of all three predictors in one regression model. We repeated these steps to explore the effect of teaching experience and subjective appraisals on changes in teachers' HR at each interval.

3. Results

3.1. 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

[5]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).

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.

“‘{=latex table_1}

Interval	*M HR*	*SD HR*	Min	Max
Overall Course of 2h	90.09/0.04[6]	15.76/0.991	51/-4.03	164/4.56
Pre-teaching interval (I_1)	96.28/0.48	14.11/0.88	56/-3.56	139/3.24
Teaching interval (I_2)	100.80/0.85	16.23/0.77	63/-2.18	164/4.37
Post-teaching interval (I_3)	93.61/0.27	14.01/0.76	60/-2.17	150/3.06
Interview interval (I_4)	82.32/-0.72	11.85/0.74	51/-2.51	132/4.39
End interval I_5)	77.95/-1.07	11.14/0.57	50[7]/-2.68	120/2.96

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).

```
‘‘‘{=latex}
\begin{figure}[H]
  \centering
  \includegraphics[width=1\textwidth]{plots_publication/loess_plot_std_unstd_new.pdf}
  \caption{Overall course of the HR with the unstandardized HR in BPM shown in Fig. 3a. and the z-star
  \label{Overall course of the HR with the unstandardized HR in BPM shown in Fig. 3a. and the z-standa
\end{figure}
```

We first tested the hypothesis that teachers showed the highest mean standardized HR during the micro-teaching unit, compared to all other intervals (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 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).

[0] Please note that standardized M and SD of the overall course were not exactly 0 and 1 due to rounding differences

[1] Deviations of the minimum values in the overall course vs. the end interval (I_5) are due to data of a few participants who needed more than two hours to finish the study.

[2] Please note that standardized M and SD of the overall course were not exactly 0 and 1 due to rounding differences

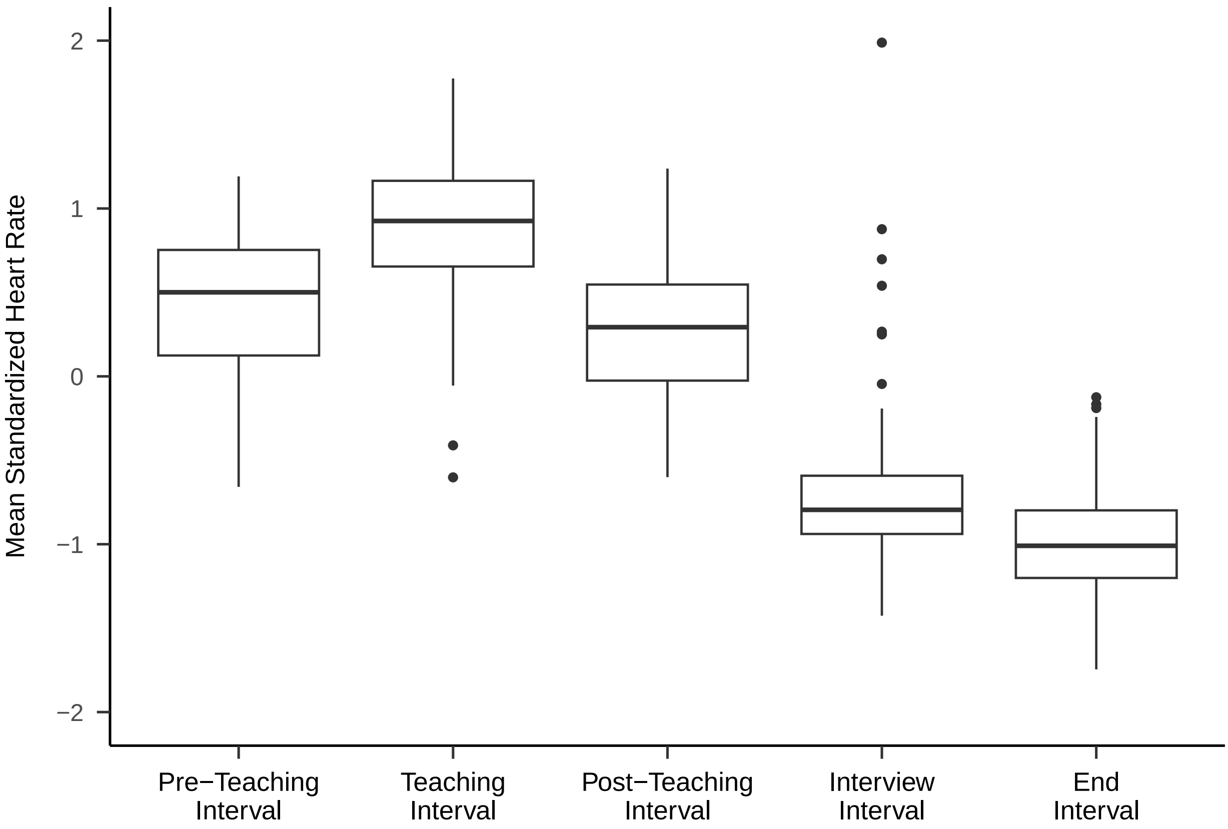
[3] Deviations of the minimum values in the overall course vs. the end interval (I_5) are due to data of a few participants who needed more than two hours to finish the study.

[4] Please note that standardized M and SD of the overall course were not exactly 0 and 1 due to rounding differences

[5] Deviations of the minimum values in the overall course vs. the end interval (I_5) are due to data of a few participants who needed more than two hours to finish the study.

[6] Please note that standardized M and SD of the overall course were not exactly 0 and 1 due to rounding differences

[7] Deviations of the minimum values in the overall course vs. the end interval (I_5) are due to data of a few participants who needed more than two hours to finish the study.



Note: N = 81 participants per interval. Fig. shows median (bold line), interquartile range (box) and outliers (dots).

Figure 3: Distribution of the standardized heart rate means in the five intervals

Next, we examined HR changes (i.e., mean slopes) within each interval to test for the hypothesis that HR increased during the pre-teaching interval and decreased during all other 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 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.

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

Correlations among mean standardized HR/mean slopes (see Table 2), teaching experience (M = 5.64, SD = 9.46), disruption appraisal (M = 5.19, SD = 2.87), and confidence appraisal (M = 7.81, SD = 1.97) are

[6] All measurement points per interval for all participants. Note that the variation in n stems from the variation in the number of collected data points by the fitness tracker.
[7] All measurement points per interval for all participants. Note that the variation in n stems from the variation in the number of collected data points by the fitness tracker.
[8] All measurement points per interval for all participants. Note that the variation in n stems from the variation in the number of collected data points by the fitness tracker.

Interval	n[8]	*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. * $p < .05$

Table 2: Descriptive statistics (n, M, SD) for the mean intercepts and the mean slopes for the five intervals.

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, 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.

Neither adding disruption appraisal (**Hypothesis 2b**) nor adding confidence appraisal (**Hypothesis 2c**) while controlling for the shared variance with teaching experience revealed 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 the other variables.

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, no variable revealed a significant prediction.

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 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.

Dependent variable:	Model 1				Model 2				Model 3				Model 4			
	Mean standardized HR and mean slopes															
	Mean std. HR		Mean slopes		Mean std. HR		Mean slopes		Mean std. HR		Mean slopes		Mean std. 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)																
Teaching Experience	−.17 (.005)	.12	−.27* (.002)	< .05												
R^2	.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								
R^2	.012		.000		.040		.015		.012		.010		.042		.031	
Δ R^2			.028		.015		.000		.010		.030		.031			

Continued on next page

Table 4 – continued from previous page

Dependent variable:	Model 1				Model 2				Model 3				Model 4			
	Mean standardized HR and 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
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								
R^2	.002		.001		.043		.020		.006		.002		.058		.023	
ΔR^2			.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								

Continued on next page

Table 4 – continued from previous page

Dependent variable:	Model 1				Model 2				Model 3				Model 4			
	Mean standardized HR and 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
Confidence Appraisal	−.02 (.050)	.85	−.13 (.007)	.29	−.04 (.052)	.76	−.16 (.007)	.20								
R^2	.058		.040		.060		.050		.058		.054		.061		.069	
ΔR^2			.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								
R^2	.002		.013		.005		.053		.012		.025		.013		.078	
ΔR^2			.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 to teaching experience as a predictor. In Model 3, solely confidence appraisal was added to teaching experience as a predictor. In Model 4, all three predictors were considered in concert.
* $p < .05$.

References

. *Fitbit Charge 4 User Manual Version 1.2*. Fitbit, Incorp., n.d. URL `https://staticcs.fitbit.com/content/assets/help/manuals/manual_charge_4_en_US.pdf`.

John Allen. Photoplethysmography and its application in clinical physiological measurement. *Physiological measurement*, 28 (3):R1, 2007.

Ariel M Aloe, Shannon M Shisler, Benjamin D Norris, Amanda B Nickerson, and Tyler W Rinker. A multivariate meta-analysis of student misbehavior and teacher burnout. *Educational research review*, 12:30–44, 2014.

Irma Battipaglia and Gaetano A. Lanza. *The Autonomic Nervous System of the Heart*, pages 1–12. Springer Berlin Heidelberg, Berlin, Heidelberg, 2015. ISBN 978-3-662-45074-1. doi: 10.1007/978-3-662-45074-1_1. URL `https://doi.org/10.1007/978-3-662-45074-1_1`.

Taryn Chalmers, Blake Anthony Hickey, Phillip Newton, Chin-Teng Lin, David Sibbritt, Craig S McLachlan, Roderick Clifton-Bligh, John Morley, and Sara Lal. Stress watch: The use of heart rate and heart rate variability to detect stress: A pilot study using smart watch wearables. *Sensors*, 22(1):151, 2021.

Roland P Chaplain. Stress and psychological distress among trainee secondary teachers in england. *Educational Psychology*, 28(2):195–209, 2008.

John W Cheng and Hitoshi Mitomo. The underlying factors of the perceived usefulness of using smart wearable devices for disaster applications. *Telematics and Informatics*, 34(2):528–539, 2017.

Francisco de Arriba-Pérez, Manuel Caeiro-Rodríguez, and Juan Manuel Santos-Gago. Towards the use of commercial wrist wearables in education. In *2017 4th Experiment@ International Conference (exp. at'17)*, pages 323–328. IEEE, 2017.

Monika H Donker, Tamara Van Gog, and M Tim Mainhard. A quantitative exploration of two teachers with contrasting emotions: intra-individual process analyses of physiology and interpersonal behavior. *Frontline Learning Research*, 6(3): 162–184, 2018.

Ty Ferguson, Alex V Rowlands, Tim Olds, and Carol Maher. The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. *International journal of behavioral nutrition and physical activity*, 12 (1):1–9, 2015.

Molly H Fisher. Factors influencing stress, burnout, and retention of secondary teachers. *Current issues in education*, 14(1), 2011.

Alan Godfrey, Victoria Hetherington, Hubert Shum, Paolo Bonato, NH Lovell, and Sam Stuart. From a to z: Wearable technology explained. *Maturitas*, 113:40–47, 2018.

T Hao, H Chang, M Ball, K Lin, and X Zhu. chrv uncovering daily stress dynamics using bio-signal from consumer wearables. In *In MEDINFO 2017: Precision Healthcare Through Informatics: Proceedings of the 16th World Congress on Medical and Health Informatics*, volume 245, page 98, 2018.

Keith C Herman, Sara L Prewett, Colleen L Eddy, Alyson Savala, and Wendy M Reinke. Profiles of middle school teacher stress and coping: Concurrent and prospective correlates. *Journal of School Psychology*, 78:54–68, 2020.

K. Hottenrot. *Trainingskontrolle: mit Herzfrequenz-Messgeräten, 2. Auflage*. Meyer & Meyer Verlag, 2007.

Yizhen Huang, Eric Richter, Thilo Kleickmann, and Dirk Richter. Class size affects preservice teachers' physiological and psychological stress reactions: An experiment in a virtual reality classroom. *Computers & Education*, 184:104503, 2022.

Andrew Hughes, Md Mobashir Hasan Shandhi, Hiral Master, Jessilyn Dunn, and Evan Brittain. Wearable devices in cardiovascular medicine. *Circulation Research*, 132(5):652–670, 2023.

Richard M Ingersoll and Thomas M Smith. The wrong solution to the teacher shortage. *Educational leadership*, 60(8):30–33, 2003.

Mary Renck Jalongo and Kelly Heider. Editorial teacher attrition: An issue of national concern. *Early Childhood Education Journal*, 33:379–380, 2006.

Robin Junker, Monika H Donker, and Tim Mainhard. Potential classroom stressors of teachers: An audiovisual and physiological approach. *Learning and Instruction*, 75:101495, 2021.

Chris Kyriacou. Teacher stress: Directions for future research. *Educational review*, 53(1):27–35, 2001.

Chris Kyriacou and James Sutcliffe. Teacher stress: Prevalence, sources, and symptoms. *British journal of educational psychology*, 48(2):159–167, 1978.

Richard S Lazarus and Susan Folkman. *Stress, appraisal, and coping*. Springer publishing company, 1984.

Richard S Lazarus and Raymond Launier. Stressbezogene transaktionen zwischen person und umwelt. *Stress. Theorien, Untersuchungen, Maßnahmen*, pages 213–259, 1981.

Meihua Liu and Yi Yan. Anxiety and stress in in-service chinese university teachers of arts. *International journal of higher education*, 9(1):237–248, 2020.

Subhas Chandra Mukhopadhyay and Tarikul Islam. *Wearable Sensors: Applications, design and implementation*. IOP Publishing, 2017.

Kayla Nuss, Kristen Moore, Tracy Nelson, and Kaigang Li. Effects of motivational interviewing and wearable fitness trackers on motivation and physical activity: A systematic review. *American Journal of Health Promotion*, 35(2):226–235, 2021.

Eunil Park. User acceptance of smart wearable devices: An expectation-confirmation model approach. *Telematics and Informatics*, 47:101318, 2020.

Chenming Peng, Nannan Xi, Hong Zhao, and Juho Hamari. Acceptance of wearable technology: a meta-analysis. In *Hawaii International Conference on System Sciences*, pages 5101–5110, 2022.

Tam Pham, Zen Juen Lau, SH Annabel Chen, and Dominique Makowski. Heart rate variability in psychology: A review of hrv indices and an analysis tutorial. *Sensors*, 21(12):3998, 2021.

Rebecca Quintana, Chris Quintana, Cheryl Madeira, and James D Slotta. Keeping watch: Exploring wearable technology designs for k-12 teachers. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 2272–2278, 2016.

RStudio Team. *RStudio: Integrated Development Environment for R.* RStudio, PBC., Boston, MA, 2020. URL http://www.rstudio.com/.

N Runge, S Haarman, and M Fisher. Using fitbit fitness trackers to measure teacher stress and coping. *International Journal of Social Policy and Education*, 2:56–70, 2020.

S Sammito, B Thielmann, R Seibt, A Klussmann, M Weippert, and I Böckelmann. Guideline for the application of heart rate and heart rate variability in occupational medicine and occupational science. *ASU Int*, 2015(06):1–29, 2015.

Klaus Scheuch and Monika Knothe. Psychophysische beanspruchung von lehrern in der unterrichtstätigkeit. *Jahrbuch für Lehrerforschung*, 1(S 285):299, 1997.

A Smith. The scale of perceived occupational stress. *Occupational Medicine*, 50(5):294–298, 2000.

M. Sperka and U.. Kittler. *Psychophysiologische Beanspruchung von Lehramtskandidatinnen und -kandidaten im Schulunterricht*, pages 182–197. Die Blaue Eule, Essen, 1995.

Joachim Taelman, Steven Vandeput, Arthur Spaepen, and Sabine Van Huffel. Influence of mental stress on heart rate and heart rate variability. In *4th European conference of the international federation for medical and biological engineering*, pages 1366–1369. Springer, 2009.

Thomas Unterbrink, Anna Hack, Ruth Pfeifer, Veronika Buhl-Grießhaber, Udo Müller, Helmut Wesche, Markus Frommhold, Klaus Scheuch, Reingard Seibt, Michael Wirsching, et al. Burnout and effort–reward-imbalance in a sample of 949 german teachers. *International archives of occupational and environmental health*, 80:433–441, 2007.

Alexander Wettstein, Sandra Schneider, Martin grosse Holtforth, and Roberto La Marca. Teacher Stress: A Psychobiological Approach to Stressful Interactions in the Classroom. *Frontiers in Education*, 6, 2021. ISSN 2504-284X. URL https://www.frontiersin.org/articles/10.3389/feduc.2021.681258.