**# Introduction**

In educational contexts, there’s a huge interest in exploring whether heart rate (HR) measures serve as reliable indicators for teachers’ stress during teaching settings [@karner2021teachers; @wettstein2020ambulatory]. Prior studies showed that teacher-centered activities and typical stressors such as classroom disruptions lead to increased teacher HRs in teaching settings. However, these studies often relied on expensive and invasive electrocardiographs (ECG) to measure teachers’ HR [@sperka1995; @scheuch1997psychophysische; @donker2018; @junker2021; @huang2022class]. Given the high stress levels in the teaching profession [@johnson2005experience], using affordable, highly accepted, and non-invasive instruments like wrist-worn fitness trackers [@ferguson2015] could be a valuable tool for analyzing teachers’ HR and the factors contributing to teachers’ stress in non-intrusive everyday teaching situations. Thus, 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.

According to Lazarus and Folkman’s (1987) transactional model of stress and coping,

As teachers are confronted with a multitude of demands in their everyday work that might exceed their resources at hand, they might have difficulties coping 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 disruptions, for example, are one of the major stressors in teachers’ daily work [@boyle1995structural; @aloe2014multivariate] 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].

Wrist-worn fitness trackers that measure teachers’ HR might contribute to a better understanding of how stressors like classroom disruptions affect teachers’ stress responses. [@wettstein2021]. Therefore, this study explored the use of wrist-based fitness trackers as a tool to assess teachers’ HR as an indicator of stress before, during, and after a teaching session in which typical classroom disruptions occurred. Teachers’ HR data were triangulated with teachers’ appraisals of classroom disruptions and their teaching experience.

**## Fitness trackers as a method to assess stress**

Wearables, defined as electronic devices worn directly on the body or integrated into clothing or accessories, serve as versatile solutions [@godfrey2018z], gathering data like location, movements, and vital signs [@cheng2017underlying]. Fitness trackers, a popular wearable technology [@park2020user], provide data on physical activity and cardiovascular parameters such as HR, supporting personalized fitness goals [@nuss2021effects] and stress management [@hao2018chrv]. Their affordability and ease of use have contributed to their widespread use in healthcare, entertainment, and fitness [sinha2019taxonomy]. But also in education, fitness trackers offer added benefits for formal and informal learning environments for both students and teachers [@de2017towards]. Yet, few studies focused on their significance for teachers. However, unlike clinical devices, fitness trackers could offer continuous and less intrusive data collection such as heart rate (HR) over time [@godfrey2018z], aligning with the increasing popularity and acceptance of wearables among the general population [@peng2022acceptance]. Furthermore, in contrast to self-report questionnaires on stress [@chaplain2008; @liu2020] that are prone to biases like social desirability [@razavi2001self] or recall errors [@van2016accuracy], ambulatory assessment methods [@trull2013ambulatory; @wettstein2020ambulatory] like HR measurements can overcome these limitations by offering objective insights into teachers’ stress levels, e.g., by monitoring teachers’ HR without disrupting teaching [@donker2018; @runge2020].

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

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

**## Teacher stress**

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 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 [@lazarus1981stressbezogene; @lazarus1984stress], which was modified and tailored to the teaching-learning environment by @kyriacou1978. In general, the transactional stress model [@lazarus1990theory] highlights the interaction between an individual and the environment, whereby stress refers to a person´s subjective reaction to an event (a stressor) that exceeds their adaptive resources.

**Figure 1**

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



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

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

Prior research, not using fitness trackers but traditional electrocardiography (ECG), has shown that changes in teacher HR can be mapped onto stressors experienced by teachers during teaching. For example, teachers’ HR increased during teacher-centered activities when they had to take an exposed position in student-teacher interaction [@sperka1995; @scheuch1997psychophysische; @donker2018; @junker2021]. @sperka1995 for example recorded the HR of 16 pre-service teachers during their first lesson. 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 finding as indicating that pre-service teachers’ proactive coping strategies, such as actively managing student interactions, helped in lowering their HR levels. 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 during a real classroom lesson. They provided evidence that teacher stress, induced by factors like low student engagement (e.g., lack of motivation or interest in tasks) or teacher-centered activities (e.g., teacher-focused classroom activities), resulted in elevated HR.

In addition to ECG studies, there are a few studies that used wrist-worn fitness trackers to investigate HR trends in teaching-learning situations [@Darnell2019; @chalmers2021]. @Darnell2019 for example measured the HR of 15 medical college students listening to a lecture using wrist-worn devices during lecture classes. 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 (peer-discussion based problem solving). @chalmers2021 examined the usability of the average HR, measured with a fitness tracker, to identify physiological changes during stress-inducing tasks with the Trier Stress Test in a study with 30 medical students and 30 participants from the general population. 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: @runge2020 used a Fitbit fitness tracker to assess HR as an indicator of stress in *N*=4 in-service teachers. They used the fitness trackers’ recordings to create a profile for each teacher, with the aim of differentiating between teachers reporting higher or lower levels of stress. In particular, it was found that the combination of a high number of steps, a high HR, and short sleep was characteristic of teachers reporting high stress levels. It should, however, be noted that the generalizability of the results is limited due to the small sample size.

**## The role of appraisals and teaching experience in teachers’ stress response**

As shown in Fig. 1., 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 [@jalongo2006; @unterbrink2007; @aloe2014]. This highlights the importance of teachers’ appraisal of their professional competencies, including their ability to manage classroom disruptions effectively [@klieme2008concept; @konig2016teacher].

As shown in Fig. 1, both primary and secondary appraisals are influenced by teachers’ characteristics such as their professional experience, shaping their classroom management skills. As professional experience grows, teachers develop cognitive scripts for managing classroom events, resulting in more complex and effective classroom management skills [@wolff2021classroom]. Especially beginning teachers face considerable stress and often feel overwhelmed by the demands of teaching [@ophardt2017klassenmanagement; @wolff2015keeping; @ klusmann2012berufliche], with many leaving the profession within the first five years [@ingersoll2003]. Research shows less experienced teachers are more susceptible to burnout, underscoring the importance of professional experience in predicting teacher stress [@fisher2011].

In summary, previous studies have revealed that teachers’ (and students’) HR changes, depending on their activities and the stressors they experience, whereby particularly teacher-centered phases 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 are still lacking, that investigate data from teacher-worn fitness trackers in larger samples to explore the links between teachers’ HR and subjective stressor appraisal or effects of teaching experience.

**## Present Study**

The data analyzed in the present study were obtained from in-service and pre-service teachers who participated in a laboratory study as part of a larger project targeting the development of classroom management.

As part of the larger project, participants came to the lab individually and taught a 15-minute, self-prepared micro-teaching unit to a “class” of three actors (trained student assistants). These actors performed nine possibly disruptive, typical 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. The micro-teaching unit was potentially stressful for the participants, given its unfamiliar setting and the scripted disruptions of participants’ teaching flow. Thus, we were particularly interested in the changes in 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 possibly disruptive classroom events. In the *post-teaching phase*, participants answered several questionnaires. Next, in the *interview phase,* participants engaged in a stimulated recall interview during which they rated for each classroom events how disruptive it had been to them, and how confident they had felt in dealing with it. In the *end phase*, participants answered another questionnaire. These sequences were identical for all participants. During the entire study, participants wore a fitness tracker on their wrist.

The aims of the present study were twofold:

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

First, we expected the participants to show an initial increase in their HR, followed by a peak during the micro-teaching unit and a decrease for the remaining phases. In addition, we examined whether z-standardization of the participants’ 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 also Figure 2: pre-teaching interval (I1), teaching interval (I2), post-teaching interval (I3), interview interval (I4), end interval (I5)). We examined HR level and change during these intervals in order to test the hypotheses that a) teachers would show the highest mean HR during the micro-teaching unit (I2), compared to all other phases (\*\*Hypothesis 1a\*\*), and b) that teacher HR would increase while they were preparing for teaching during the pre-teaching interval (I1), but decrease in all of the following intervals, because of habituating to (I2) and recovering from (I3-I5) the potentially stressful micro-teaching unit (\*\*Hypothesis 1b\*\*).

(2) We further explored whether teaching experience would make a difference in how teachers reacted to the classroom disruptions. In line with the research on teacher expertise and teacher stress reviewed above, we expected more experienced teachers might have better classroom management strategies, and thus better resources for coping. In addition, we were interested in finding a relationship between teachers’ appraisals of the classroom events (disruptiveness; confidence in dealing with them) and teachers’ HR, beyond the explanatory power of teaching experience. Specifically, we expected lower HR levels for teachers with more teaching experience, particularly during the micro-teaching unit (\*\*Hypothesis 2a). We expected higher HR levels for teachers who felt more disrupted, 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 analogous hypotheses for the *changes* in HR.

**# Method**

**## Participants**

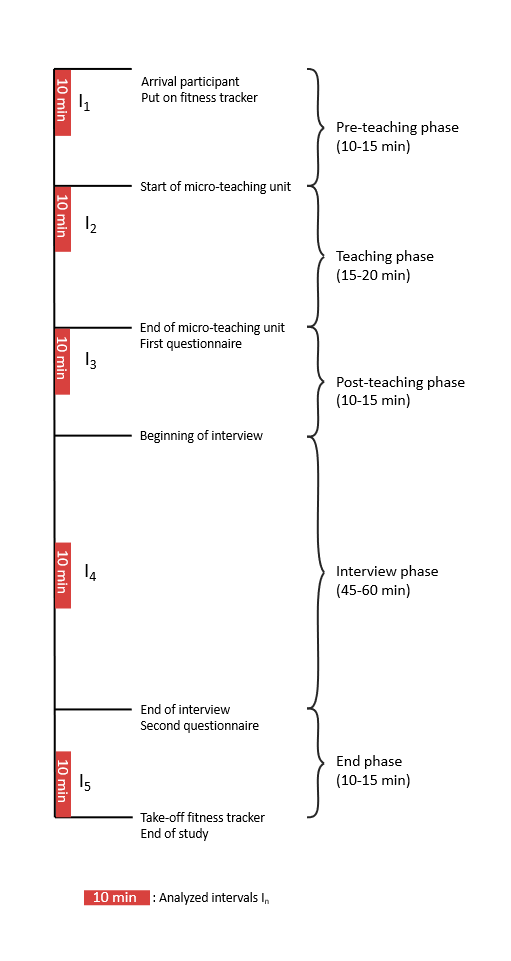
The sample consisted of *N* = 84 pre- and in-service teachers from Germany, who were recruited via personal contact, email lists, and flyers. The data of three participants was lost due to failed data transmission, yielding an analysis sample of *n*total= 81 (*n*total = 52 women, *n*total = 29 men),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

**## Setting and Procedure**

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

**Figure 2**

*Procedure of the two-hour-long study, consisting of five phases with five representative 10-minute intervals*

**

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

**## Measures**

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

To measure teachers’ HR, we used the wrist-based fitness tracker Fitbit Charge 4. In line with the manufacturer's instructions [@fitbitnd], the device was attached 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, to calculate the 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[[1]](#footnote-1). The raw data that can be extracted from the tracker contains the time stamps of all measurements and the estimated HR in BPM for each time stamp. 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 and to keep measurement intervals comparable within phases, we aggregated HR over a representative 10-minute interval within each phase. Previous research has indicated that 10-minute intervals are a useful duration for analyzing PPG data [@lu2008can]. The intervals were selected based on the following rules: The pre-teaching interval (I1) comprised the first 10 minutes after the fitness tracker had been put on. The teaching interval (I2) started two minutes after the teacher had started the teaching unit. This interval was of the highest relevance to our study. We explicitly chose an early 10-minute interval within the *teaching phase*, as previous studies revealed that the beginning of a lesson is essential and demanding regarding teacher-student interaction [@donker2018quantitative; @claessens2017positive]. The post-teaching interval (I3) started immediately after the end of the teaching unit. The interview interval (I4) was defined as the mid-10 minutes between the end of the teaching unit and the time point when the fitness tracker was taken off so that all participants were being interviewed during this interval. The end interval (I5) comprised the last 10 minutes before the fitness tracker was taken off.

**### Teaching Experience**

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

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

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

**## Data analysis**

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

\*\*Research goal 1\*\*. The first research goal included mapping teachers’ HR before, during, and following a micro-teaching unit over 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. To account for individual differences in the baseline HR, we visually compared unstandardized and standardized mean HR trends over the course of the entire two-hour study.

For testing Hypothesis 1a, which examined mean differences of the HR levels across the selected five intervals, we initially conducted a one-way ANOVA with repeated measures as an omnibus test. The dependent variable 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 standardized 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 standardized HR values over time. To this end, we used fixed intercept fixed slope regression models [@gelman2006data] for each interval to estimate intercepts and linear slopes for all individuals which were then averaged across individuals.[[2]](#footnote-2)

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

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

**# 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’ HR is a useful method to account for individual differences in the baseline HR. Means, standard deviations, and range of teachers’ unstandardized and standardized HR are shown in Table 1. Fig. 3 a. and b. display the unstandardized and standardized HR, respectively.  HR initially increased, peaked, and then decreased, with the unstandardized and standardized HR graphs showing high similarity. Thus, for all further analyses, we used participants’ standardized HR values.

**Table 1**

*Mean HR (*M*), standard deviations HR (*SD*), and Range of Teachers’ HR Over the Course of the Entire Study and the Five Intervals (Unstandardized in BPM/z-standardized)*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Interval | *M HR* | *SD HR* | Min | Max | |
| Overall Course | 90.09/0.041 | 15.76/0.991 | 51/-4.03 | 164/4.56 | |
| Pre-Teaching Interval (I1) | 96.28/0.48 | 14.11/0.88 | 56/-3.56 | 139/3.24 | |
| Teaching Interval (I2) | 100.80/0.85 | 16.23/0.77 | 63/-2.18 | 164/4.37 | |
| Post-Teaching Interval (I3) | 93.61/0.27 | 14.01/0.76 | 60/-2.17 | 150/3.06 | |
| Interview Interval (I4) | 82.32/-0.72 | 11.85/0.74 | 51/-2.51 | 132/4.39 | |
| End Interval (I5) | 77.95/-1.07 | 11.14/0.57 | 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.*

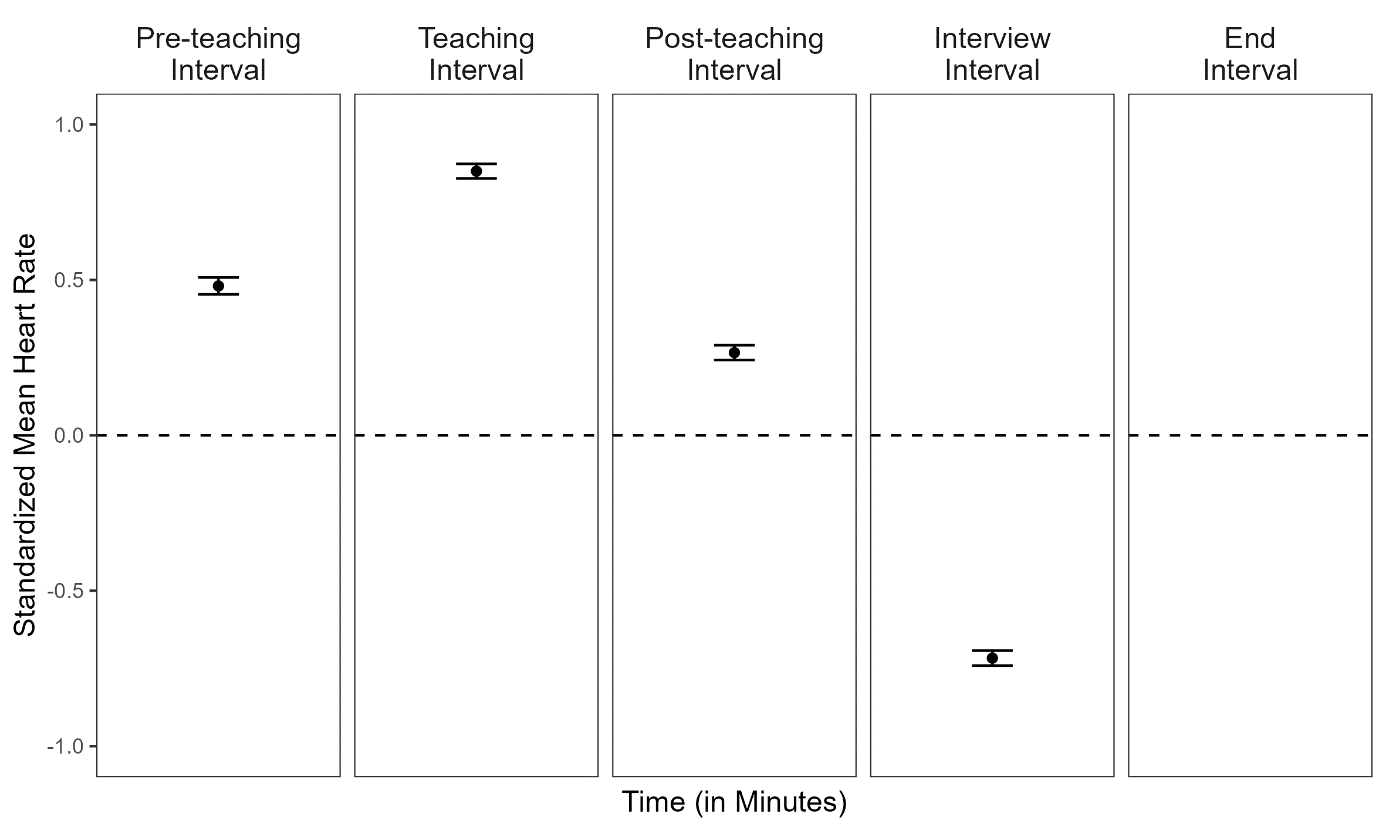


*Note:* The shadow around the line represents the 95% confidence interval. The confidence interval shown refers to the HR measurement points during the entire study period.

We first tested the hypothesis that teachers showed the highest mean HR peaked during the micro-teaching unit, compared to all other phases (Hypothesis 1a). Repeated measures ANOVA revealed that the standardized mean HR differed statistically significantly between intervals, *F*(4, 400) = 257.50, *p* < .05, *f* = 1.60 (large effect). Post-hoc contrasts indicated that, as hypothesized, the standardized mean HR was significantly higher in the teaching interval (I2) than in all other phases (see also Fig. 4). Specifically, it was higher than in the pre-teaching interval (I1; *t*(1) = 32.71, *p* < .05, *d* = 0.82; large effect), 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).

**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 HR changes within each interval to test for the hypothesis that HR would increase in the pre-teaching (I1) and decrease in the post-teaching (I3), interview (I4), and end intervals (I5; 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 mean of the pre-teaching interval (I1) was significantly positive, indicating an increase in HR, as hypothesized. Further, the slope means 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 slope mean was negative but did not differ significantly from zero.

**Table 2**

*Descriptive statistics* *(*n, M, SD*)* *for the mean intercepts and the mean slopes for the 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 mean slopes with teaching experience and self-report data**

Correlations among standardized mean HR/mean slopes, teaching experience, disruption appraisal, and confidence appraisal are presented in Table 3. Correlations between standardized mean 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., standardized mean HR) for each of the five intervals (Hypotheses 2a-d), teaching experience significantly predicted standardized mean HR only in the interview interval (Table 4, Interview Interval, Model 1), indicating a higher standardized mean HR for teachers with more teaching experience. This relationship is, in fact, in the opposite direction predicted by Hypothesis 2a.

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

When considering the effects of the three predictors in concert (\*\*Hypothesis 2d\*\*), standardized mean 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 standardized mean HR for teachers who felt more disrupted by the classroom events, when controlling for all other factors.

Concerning the explorative investigation of the effects of teaching experience and subjective appraisals on *changes* (i.e., mean slopes) in teachers’ HR, teaching experience significantly predicted the mean slope in the pre-teaching interval(Table 4, Pre-Teaching Interval, Model 1), indicating a less steep HR increase in teachers with more teaching experience. For all other intervals, the prediction was not significant (*p* > .05).

**Table 3**

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

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | Pre-Teaching Interval | Teaching  Interval | Post-Teaching  Interval | Interview  Interval | End  Interval |
| TE | − .17/− .27\* | .11/−.02 | − .04/−.03 | .24\*/−.20 | .04/.11 |
| DA | − .01/.16 | − .20/.08 | .20/−.14 | − .13/.01 | .04/.12 |
| CA | − .10/− .18 | .06/.09 | .04/−.03 | .09/−.19 | − .07/.13 |
| *Note.* TE = teaching experience, DA = disruption appraisal, CA = confidence appraisal,  \* *p* < .05. | | | | | |

**Table 4**

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

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

**# Discussion**

In educational contexts, recent studies aimed to find adequate indicators for teacher stress such as HR. Nevertheless, most studies used expensive and intrusive devices such as ECG devices. Fitness trackers have enormous potential to offer teachers the opportunity to use low-cost, easy-to-use, and lightweight devices to record HR as an indicator of stress during teaching. Up to now, there has been a gap in research on the extent to which fitness trackers are effective and suitable for recording HR measurements during teaching. Our study is the first to investigate how data collected from a wrist-worn fitness tracker could shed light on the effects of stressors, such as classroom disruptions, on teachers’ stress responses before, during, and after teaching sessions. Therefore, we measured teachers’ HR over the course of a five-phase lab study with a comparatively large sample. Moreover, we examined whether variance in HR measures can be explained by teachers’ teaching experience and self-reported appraisal (disruption and confidence appraisal) of potentially disruptive classroom events. Building on the model of teacher stress [@kyriacou1978], we hypothesized that …

## **Mapping HR over study phases**

Our first research question concerned the effectiveness and suitability of HR measures assessed by wrist-based fitness trackers for mapping teachers’ HR over the course of the five-phase lab study, including the time before, during, and after the potentially stressful micro-teaching unit. As expected, teachers’ HR increased before the micro-teaching unit, peaked during, and progressively decreased after the teaching phase. Furthermore, we investigated whether z-standardization of the participants’ mean HR could serve as a useful method to account for individual differences in baseline HR. Results revealed that the standardized and non-standardized mean HR values were comparable, i.e., this method ensured that any observed differences in HR between individuals were not solely due to inherent differences in baseline HR levels (see ##Limitations).

Our \*\*Hypothesis 1a\*\*, according to which teachers would show the highest mean HR during the micro-teaching unit (I2), compared to all other phases, was confirmed. This is in line with prior studies showing that teachers’ HR varies depending on their activities and encountered stressors, particularly increasing during phases where teachers are in an exposed position [@sperka1995; @scheuch1997psychophysische; @donker2018]. In comparison to the other phases of our study, the micro-teaching unit was the phase in which the participants reached the highest HR and therefore also possibly the highest stress level. This can be explained by the fact that the micro-teaching unit was the phase in which teachers were the center of attention because they had to actively teach in front of a class consisting of three actors. Teacher-centered activities, historically overlooked in studies on psychological stress, emerged as the most influential factor in inducing stress among educators, with important implications for their psychological and physiological well-being [@junker2021].

Regarding our \*\*Hypothesis 1b\*\*, according to which teachers’ HR would increase during the pre-teaching interval (I1) but subsequently decline in the subsequent intervals (I3-I5), was confirmed as well. Our findings suggest that teachers’ HR increased due to the preparation for the micro-teaching unit and decreased as a process of habituation to the teaching environment and recovery from the potentially stressful micro-teaching unit.

visualizations and analyses of the HR data in the five different phases and throughout the entire study clearly showed that wrist-worn fitness trackers are a suitable method for measuring different states of arousal in (prospective) teachers in a laboratory setting.

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

(2) We further explored whether teaching experience would make a difference in how teachers reacted to the classroom disruptions. In line with the research on teacher expertise and teacher stress reviewed above, we expected more experienced teachers might have better classroom management strategies, and thus better resources for coping. In addition, we were interested in finding a relationship between teachers’ appraisals of the classroom events (disruptiveness; confidence in dealing with them) and teachers’ HR, beyond the explanatory power of teaching experience. Specifically, we expected lower HR levels for teachers with more teaching experience, particularly during the micro-teaching unit (\*\*Hypothesis 2a). We expected higher HR levels for teachers who felt more disrupted, 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 analogous hypotheses for the *changes* in HR.

## **Prediction of standardized mean HR and mean slopes with teaching experience and subjective appraisals**

Regarding our second research question, the linear regression models only partially supported our expectations.

**## Limitations**

As with all research, methodological limitations need to be considered.

Firstly, the laboratory setting of the study did not correspond to an authentic classroom environment. The data collected during the micro-teaching unit in a highly standardized situation was based on a fictional setting, resulting in various aspects, such as the lack of a teacher-student relationship. Nevertheless, researchers agree that positive relationships are the basis for effective classroom management and productive handling of classroom disruptions [@ruedi2014; @beaty2010]. For this reason, it cannot be excluded that the artificial laboratory situation led to an increased HR of the subjects due to nervousness and excitement. However, these conditions were identical for all participants, meaning that

Secondly, the subjects’ resting HR was not recorded during the study, although this is an important marker to consider inter- and intrapersonal cardiovascular differences in participants [@nanchen2018; @heneghan2019]. For this purpose, it is necessary to record the HR during a resting phase without physical or emotional stress, ideally fifteen minutes before the beginning of the activity, to determine a valid baseline HR. This is a necessary condition when carrying out tests in the laboratory, but in practice, it can be a challenge to achieve these conditions [@sammito2015guideline] due to time constraints and difficulties in the acquisition of participants. To account for individual differences in the baseline HR without measuring the resting HR for 15 minutes, we z-standardized the BPM values from participants’ mean HRs (see ## in the Method Section).

Third, previous research has discussed the limitations of self-report data due to methodological bias [@pintrich2003; @parker2012]. pintrich2003 suggests the collection of other measures such as physiological or behavioral data in addition to self-report data to build a stronger empirical base. However, self-report data can be useful to capture intra-psychic constructs as they are in the case of this research.

## Practical implications and hands-on tips with fitness trackers

The results of our study point to several important implications.

The increasing availability of HR data through wearable fitness trackers offers consumers the opportunity to self-monitor important mental health indicators such as HR as an indicator of stress, beyond traditional self-reporting or expensive, intrusive ECG devices. Using fitness trackers could enable teachers to strengthen their self-awareness in stressful situations and allow for early self-intervention. Research on mental health has shown that to achieve a regular and meaningful use of fitness trackers as mental health monitoring, it would be essential for the subjects to understand the purpose of using the fitness watches as well as the meaning of the data [@ng2018].

Important aspects to consider for future research when using wrist-worn fitness trackers include, noting the time to make the HR measurable at different intervals, e.g., to replicate findings on whether HR differs in teacher-centered and student-centered activities using fitness trackers [@junker2021]. Another point would be the collection of further external factors (social, personal, psychological, environmental, and behavioral) that influence HR, e.g., by questionnaires, to be able to make statements about the individual health and fitness status of the participants [@wang2022].

Future research could use low-cost and non-invasive devices to accompany teachers in their everyday school practice to gain insight into teachers’ stress experience during teaching. Even in teacher training, wearable fitness trackers during internships could provide new insights into the stress experience of student teachers. Thus, the evaluation of the data with the participants could provide clues as to which moments in the teaching profession and teacher training are experienced as particularly stressful to discuss and implement possible stress-reducing measures in teacher training.

**## Conclusion**

**APPENDIX**

**Figure XX**

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

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

Automatisch generierte Beschreibung

**Figure XX**

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

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

Automatisch generierte Beschreibung

**Figure XX**











1. The fluctuations in the number of seconds in which the HR was measured are due to the participants' movements, meaning that the device could not measure the HR every second. [↑](#footnote-ref-1)
2. 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-2)