***Working with this manuscript:::***

The methods/analysis/results section and supplements of the paper will be written in Rmarkdown which allows the publishing of a fully reproducible manuscript (data and analysis code). In order to write the intro and discussion together in google docs in a way that can be quickly merged with the manuscript, here are some things to consider while writing:

* ***General****:*
  + note that formatting like bold or italics will not be preserved in the markdown manuscript unless its marked with the respective syntax:
    - one \* for italics: \*example\* → *example*
    - two times \* for bold print: \*\*example\*\* → **example**
* ***Headlines****:*
  + All headlines are marked with # (# heading)
  + smaller subheadings are created by adding # (## subheading level 2; ### subheading level 3)
* ***Citations***
  + if you are in a hurry, just past the DOI of the respective paper in brackets. I’ll sort it out later.
  + If you want to be a real sweetheart and add a full citation:
    - go to googlescholar and search the paper
    - click on the citation icon and then select bibtex
    - copy the entire bibtex entry into the googledoc named “manuscript\_bibtex”. it's in the same google-folder as this document
    - copy the citation ID from the bibtex-code (that would be the entry after the first curly bracket and before the first comma. Its usually firstauthor+year+firstwordtitle (e.g. tomasello2010origins)
    - add the citation in text by using the following syntax: [@paperID]
      * e.g....as discussed in previous work [@tomasello2010origins]

**Thanks for your patience!**

**## Present Investigation**

The present investigation was part of a larger project targeting the development of professional vision in teachers. The study was carried out in a classroom at the university that served as the lab.

Within the time frame of approximately two hours, we distinguished five phases of our study: In the (1) pre-teaching phase, the participants were welcomed, prepared for the following micro teaching unit, and familiarized with the setting. During the (2) teaching phase, the participants taught a 15-minute self-prepared micro teaching unit to a "class" of three actors that performed nine (possibly disruptive) classroom events (e.g., chatting with neighbor, heckling, looking at phone; see Table ## in supplementary material for an overview and categorization of all events; also see Fig## for a depiction of the laboratory setting of the micro-teaching unit). The actors received standardized instructions on a screen (only visible to the actors, but not to the participants) to perform a classroom event every one and a half minutes.[[1]](#footnote-1) While teaching, participants wore eye-tracking glasses and additionally, their lessons were recorded by cameras. In the (3) post-teaching phase, the participants answered several questionnaires, followed by the (4) interview phase, in which they watched the video of their 15-minute unit and answered questions about the (disruptive) classroom events. In the (5) end phase, the participant answered another questionnaire. These conditions were identical for all participants.

During the entire study, the participants wore a fitness tracker, while the HR measurements provided the database for the present investigation. Previous research found that fitness trackers can be used as a low-cost, non-invasive method of measuring HR [hajj2022wrist; @fuller2020reliability] and that fitness trackers can help to detect differences in mean HR between different teaching phases [@donker2020associations; @junker2021potential]. However, XXX

Thus, our study had two research questions:

1. The first research question 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 a micro teaching unit. Therefore, we exploratively described the participants’ overall HR trend during the two-hours study interval, and to examine whether z-standardization of the participants’ mean HR can serve as a useful method to account for individual differences in baseline HR. We expected participants´ HR to gradually increase during the pre-teaching phase, to peak during the teaching phase, and to gradually decrease during the remaining time interval. Furthermore, we expected to observe the same trends in both standardized and non-standardized mean HR values.
2. The second 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 a micro teaching unit. We presumed the highest HR levels in the micro teaching unit and regarding HR changes, we expected an increase in the beginning of the study and a decrease in the following phases.
3. The second research question was to examine whether variance in HR measures can be explained by teachers’ teaching experience, and by self-reported cognitive appraisal of classroom events. We expected all three variables to be significant predictors for the HR measurements in the different phases.

**# Method**

**## Participants**

The sample consisted of *N* = 84 pre- and in-service teachers from Germany who had been recruited via personal contact, email lists, and flyers. The data of three participants was lost due to failed data transmission, yielding an analysis sample of *n* = 81.

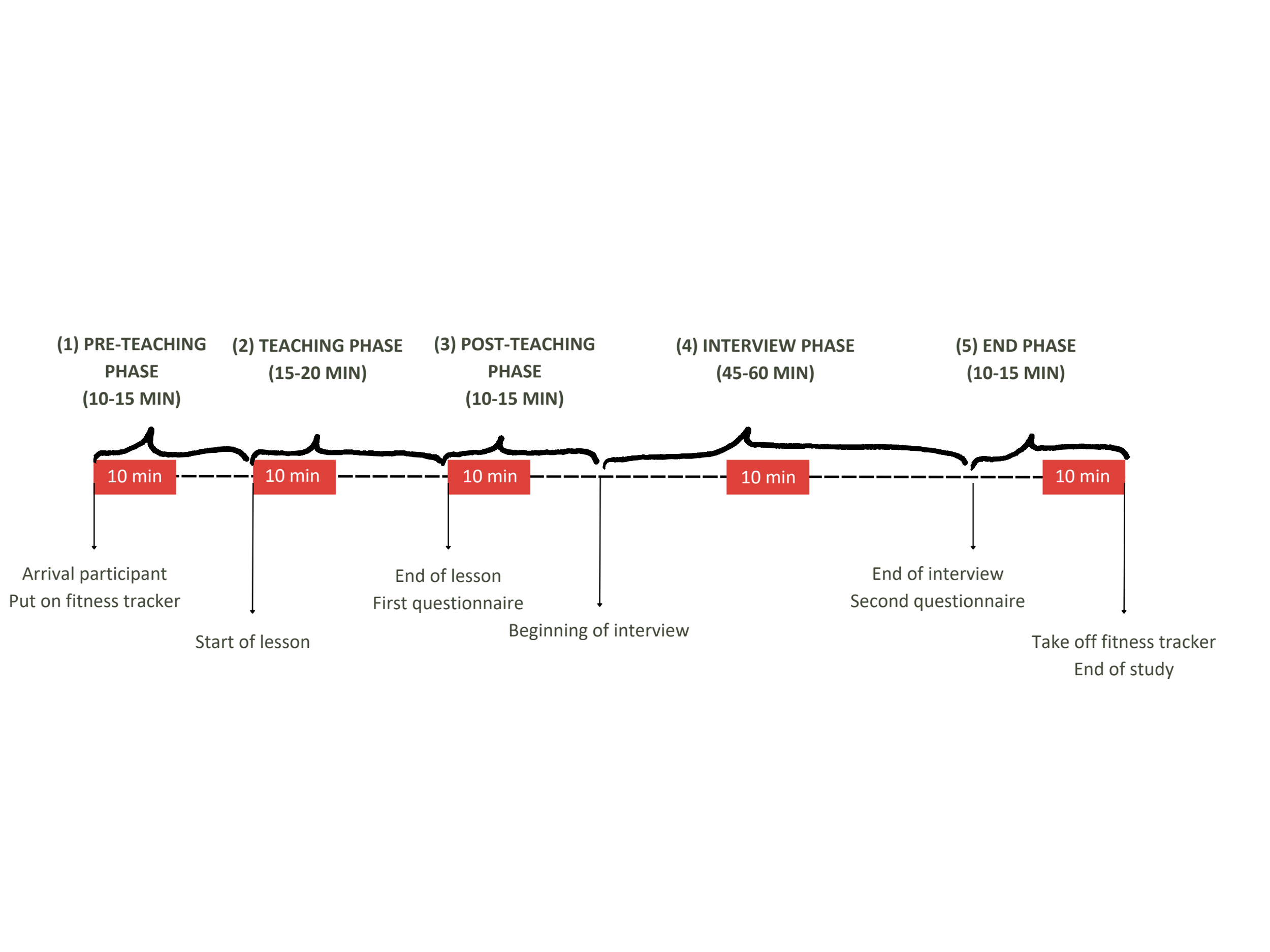
The participants of the analysis sample (*n* = 52 women, *n* = 29 men) reported a mean age of 30.95 years (*SD* = 10.90; range: 19-60) and an average teaching experience of 5.64 years (*SD* = 9.46; range: 0-37).

**## Setting and Procedure**

The study was carried out in accordance with 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 prior to testing. Participation was voluntary and only took place after written consent had been given.

**Figure 1**

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

**

The whole study had a duration of approximately two hours and consisted of five phases: (1) pre-teaching phase, (2) teaching phase, (3) post-teaching phase, (4) interview phase, and (5) end phase (please refer to Fig. 1 for a timeline). In the (1) *pre-teaching phase*, the experimenter welcomed the participants and helped them to 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 (2) *teaching phase*, the participants held their self-prepared micro teaching unit to a class of three trained actors that performed nine classroom events. The teaching unit was video-recorded and lasted about 15-20 minutes. Participants spent this time mostly standing or slowly walking around. After having completed the micro-teaching unit, in the (3) *post-teaching phase*, participants were seated at a desk and filled in questionnaires. Moreover, in the post-teaching phase, the participants were given the opportunity to take a break after teaching, for example to use the restroom, drink or rest. This phase took approximately 10-15 minutes. In the (4) interview phase, the participants then watched the video of their own teaching together with the experimenter. While doing so, they were given a Stimulated Recall Interview (SRI), during which they answered questions about their cognitive appraisal of the classroom events (see instrument section; also see Fig## in the supplementary material for a depiction of the interview setting). The interview lasted about 45-60 minutes and the participants’ position was seated. The (5) end phase lasted about 10-15 minutes and participants answered in a seated position another questionnaire irrelevant to this study.

**## Instruments**

**### Fitness Tracker**

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 [@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 catching the light that is reflected back, 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[[2]](#footnote-2). The raw data that can be extracted from the tracker lists the time stamps of all measurements and the estimated HR in beats per minute (BPM) for each time stamp.

**### Questionnaire**

In the *post-teaching phase*, the teachers answered questionnaires: a brief computer-based questionnaire assessing sociodemographic data (e.g., teaching experience, gender, studied school type, studied school subjects, extracurricular teaching activities), and a short knowledge test that is irrelevant to the present study. All in all, completion of the questionnaires took about 10 minutes.

**### Stimulated Recall Interview**

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

**## Variables**

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

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

Since our aim was to explore teachers’ HR between study phases, we decided to aggregate HR over a typical interval within each phase. To keep intervals comparable in duration, we selected intervals with a length of 10 minutes each. Previous research has indicated that 10-minute intervals are a useful duration for analyzing PPG data [@lu2008can]. The intervals were selected based on the following rules: The (1) pre-teaching interval comprised the first 10 minutes after the fitness tracker had been put on. The (2) teaching interval started two minutes after the teacher had started the teaching unit. This interval was of highest relevance to our study. We explicitly chose an early 10-minutes interval within the teaching phase, as previous studies revealed that the beginning of a lesson is essential and demanding regarding teacher-student interaction [@donker2018quantitative; @claessens2017positive]. The (3) post-teaching interval started immediately after the end of the teaching unit. The (4) interview interval was defined as the mid-10 minutes between the end of the teaching unit and the time point where the fitness tracker was taken off. This definition ensured that all participants were being interviewed during this interval. The (5) end interval comprised the last 10 minutes before the fitness tracker was taken off.

**### Teaching Experience**

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

**### Cognitive appraisal of the events**

The subjective disruption and confidence appraisals assessed in the SRI on a 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*.

**## Data analysis**

We conducted all analysis with R [@RStudio2020].

\*\*Research goal 1\*\*. The first research goal referred to the teachers’ HR trend over the course of the entire study. We expected the participants to show an initial increase in their HR, followed by a peak during the *teaching phase* and a decrease over the course of the remaining phases. Therefore, we displayed the HR trend over the course of the entire study.

In a second step, we wanted to examine whether z-standardization of the participants’ mean HR is a useful method to account for individual differences in the baseline HR. We expected trends of standardized mean HR values to show a comparable course to the non-standardized mean HR values. To this end, we visually compared the unstandardized and standardized HR trend.

\*\*Research goal 2\*\*. To accomplish our second research goal, which examined the HR levels and changes during the different phases, we first selected five corresponding intervals with a length of ten minutes each out of the five phases and examined the levels of and the changes in HR of the five intervals separately. Referring to the HR levels, we presumed the highest HR level in the *teaching interval* and lower levels in all other intervals, because the level of arousal should be highest while teaching (\*\*Hypothesis 1a\*\*). For testing Hypothesis 1a, we initially conducted a one-way ANOVA with repeated measures as an omnibus test. The dependent variable comprised the standardized HR mean for each interval. In order 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 (2) teaching interval and the other four intervals.

Unterschiedliche ns aus Tabelle 1 + 2 nochmal genau erklären – wo kommen die ns her.

Regarding the HR changes, we expected an increase during the (1) pre-teaching interval as the participants’ arousal might increase in preparation of the teaching unit and a decrease in the following intervals, because of habituating to the situation (\*\*Hypothesis 1b\*\*). 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 [@gelman2006data] for each interval to estimate intercepts and linear slopes for all individuals which were then averaged across individuals.[[3]](#footnote-3)

\*\*Research goal 3\*\*. In addressing our third research goal, we examined the effects of teaching experience and cognitive appraisal of disruptive classroom events on teachers’ HR during the five phases.

First, we considered teaching experience. We expected lower HR levels and less steep HR changes for teachers with more teaching experience (\*\*Hypothesis 2a\*\*). Therefore, we investigated the effect of solely teaching experience on the participants’ HR levels and HR changes for each of the five intervals by linear regression models (Hypotheses 2a).

Second, we considered cognitive appraisal. We expected higher HR levels and steeper HR changes for teachers who felt more disrupted by the events (\*\*Hypotheses 2b\*\*), but lower HR levels and less steep HR changes for teachers who felt more confident in dealing with the events (\*\*Hypothesis 2c\*\*). Therefore, we separately augmented the models by either the disruption appraisal of the events (Hypothesis 2b) or by the confidence appraisal of dealing with the events (Hypothesis 2c), while controlling for the shared variance with teaching experience.[[4]](#footnote-4)

Lastly, we considered all three predictors (teaching experience, disruption appraisal, confidence appraisal) in concert and expected them to remain substantial predictors (\*\*Hypothesis 2d\*\*). Therefore, we examined the effects of the three predictors in concert (Hypothesis 2d).

**# Results**

**## Research goal 1: HR Course and Standardized Mean HR**

The first research goal aimed to explore whether z-standardization of the participants’ mean HR is a useful method to account for individual differences in the baseline HR. Means, standard deviations, and range of teachers’ unstandardized and standardized HR are shown in Table 1. Fig. 2 a. and b. displays the unstandardized mean HR in BPM and the standardized mean HR, respectively. Referring to the participants’ overall HR trend, HR initially increased, peaked, and then decreased. Comparing the unstandardized and standardized HR trends revealed a high similarity of the overall courses.

**Table 1**

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

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

**Figure 2**

*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 used the ggplot2 package (v3.3.3; Wickham, 2016) to calculate the moving average of the course.

**## Research Goal 2: HR Levels and HR Changes in the Five Intervals**

Hypothesis 1a aimed to locate the HR peak. Repeated measures ANOVA revealed that the standardized HR means of the intervals differed statistically significant between intervals, *F*(4, 400) = 257.50, *p* < .05, *f* = 1.60 (large effect). Post-hoc contrasts indicated that the standardized mean HR was significantly higher in the (2) teaching interval compared to the (1) pre-teaching interval, *t*(1) = 32.71, *p* < .05, *d* = 0.82 (large effect). Moreover, the standardized HR mean of the (2) teaching interval was significantly higher than in the (3) post-teaching interval, *t*(1) = 32.00, *p* < .05, *d* = 1.34 (large effect), the (4) interview interval, *t*(1) = 453.47, *p* < .05, *d* = 3.37 (large effect), and the (5) end interval, *t*(1) = 511.89, *p* < .05, *d* = 4.68 (large effect). Thus, as hypothesized, HR peaked in the (2) teaching interval (see Fig. 3).

**Figure 3**

*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.

Hypothesis 1b, examined the HR changes within each interval. 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 4. The slope means of the (1) pre-teaching interval was significantly positive, indicating a rising HR for this interval. In contrast, the slope means of the (2) teaching interval and (3) post-teaching interval were significantly negative, indicating a decreasing HR. For the last two intervals, the (4) interview interval and (5) end interval, the slope mean was also negative but did not differ significantly from zero.

**Table 2**

*Descriptive Statistics* *(*n, M, SD*)* *for the Mean Intercepts and the Mean Slopes for the Different Intervals for all Individuals*

|  |  |  |  |
| --- | --- | --- | --- |
| Interval | n1 | *M* | SD |
| (1) Pre-teaching interval | 6896 | 0.052/0.085\* | 0.820/0.133 |
| (2) Teaching interval | 7150 | 1.025\*/-0.039\* | 0.690/0.108 |
| (3) Post-teaching interval | 6664 | 0.549\*/-0.060\* | 0.547/0.101 |
| (4) Interview interval | 6287 | -0.617\*/-0.022 | 0.614/0.070 |
| (5) End interval | 5990 | -1.004\*/-0.012 | 0.500/0.074 |
| *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 4**

*Graphical Display of the Mean Slopes of the Standardized Mean HR for Each Interval*



**## Research Goal 3: Prediction of Standardized Mean HR and Slopes With Teaching Experience and Self-Report Data**

Correlations among HR, teaching experience, disruption appraisal, and confidence appraisal are presented separately for the five intervals in Table 3. Correlations between HR and the other constructs were mostly very small and statistically non-significant. Correlations among teaching experience and appraisals were substantial and in expected directions. Regarding \*\*Hypothesis 2a\*\*, teaching experience significantly predicted the participants’ mean HR only in the (4) interview interval (*b* = .012, *p* < .05, Table 4, Interview Interval) and the participants’ HR changes only in the (1) pre-teaching interval (*b* = -.004, *p* < .05, Table 4, Pre-Teaching Interval). Adding the disruption appraisal while controlling for the shared variance with teaching experience (\*\*Hypothesis 2b\*\*) revealed a significant effect for participants’ HR changes for the teaching experience as a predictor in the (4) interview interval (*b* = -.001, *p* < .05, Table 4, Interview Interval). Adding the confidence appraisal while controlling for the shared variance with teaching experience (\*\*Hypothesis 2c\*\*), teaching experience significantly predicted the participants’ mean HR in the interview interval *(b* = .013, *p* < .05, Table 4, Interview Interval, Model 3). When considering the effects of the three predictors in concert (\*\*Hypothesis 2d\*\*), mean HR was significantly predicted by the disruption appraisal in the (3) post-teaching interval (*b* = 0.084, *p* < .05, Table 4, Post-Teaching Interval, Model 4). Furthermore, HR changes were significantly predicted by disruption appraisal as well in the (5) end interval (*b* = .015, *p* < .05, Table 3, End Interval, Model 4).

**Table 3**

*Correlations Between Standardized Mean HR and the Predictor Variables Teaching Experience, Disruption Appraisal, and Confidence Appraisal for the Five Intervals*

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

**Table 4**

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

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

**APPENDIX**

**Figure XX**

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

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

Automatisch generierte Beschreibung

**Figure XX**

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

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

Automatisch generierte Beschreibung

**Figure XX**











1. To avoid sequency effects of the order of the events and the performing actors, we used a fully balanced Latin square design. [↑](#footnote-ref-1)
2. 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-2)
3. Although this procedure does not account for nonmonotonic progressions in individual HR, a graphical evaluation revealed that the linear estimates corresponded well to the majority of the cases (see XX in the supplementary material). [↑](#footnote-ref-3)
4. In order to avoid a post-diction, the HR levels and HR changes were only predicted in the (2) teaching interval, the (3) post-teaching interval, the (4) interview interval and the (5) end interval with the disruption and confidence appraisal, i.e., not in the (1) pre-teaching interval. [↑](#footnote-ref-4)