***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 aim was to investigate whether HR measures assessed by wrist-based fitness trackers are a suitable and effective method \*\*(1)\*\* to map teachers’ HR over the course of a five phase lab study, including a micro teaching unit,  and \*\*(2)\*\*  to examine whether HR measures can be predicted by teachers’ teaching experience and self-reported data on cognitive appraisal during the micro teaching unit.

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 unit to a "class" of three actors that performed nine (possibly disruptive) classroom events. 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) events. In the (5) end phase, the participant answered another questionnaire.

According to previous findings that fitness trackers can be used as a low-cost, non-invasive method of measuring HR [hajj2022wrist; @fuller2020reliability] and that different HRs of teachers can be measured in different teaching phases [@donker2020associations; @junker2021potential], we formulate the hypotheses as follows:

\*\*Hypothesis 1\*\*. In the first step, we exploratively examined the 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 (\*\*Hypothesis 1a\*\*). Doing so, the trends of the non-standardized versus standardized HR values should show a comparable course. In the second step, we 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 (2) teaching interval and lower levels in all other intervals, because the level of arousal should be highest while teaching (\*\*Hypothesis 1b\*\*). 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 1c\*\*).

\*\*Hypothesis 2\*\*. We examined the effects of teaching experience and cognitive appraisals about disruptive events that took place during the micro teaching unit on teachers’ HR. First, we considered teaching experience and expected lower HRs in teachers with more teaching experience (\*\*Hypothesis 2a\*\*). Second, we considered cognitive appraisal and expected higher HRs for teachers who felt more disrupted by the events (\*\*Hypotheses 2b\*\*), but lower HRs for teachers who felt more confident in dealing with the (disruptive) events (\*\*Hypothesis 2c\*\*). Lastly, we considered all three predictors in concert and expected them to remain substantial predictors (\*\*Hypothesis 2d\*\*).

**# Method**

**## Participants**

The sample consisted of *N* = 81 pre- and in-service teachers from Germany. The participants were recruited via personal contact, email lists, and flyers. From the originally assessed data of 84 participants, the data of three participants was excluded due to insufficient data quality, yielding the analysis sample of 81.

The participants of the analysis sample (*n* = 52 women, *n* = 29 man) 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). A percentage of 14.81/70.37/9.88% of the participants stemmed from the school forms of primary/secondary/special educational needs schools.

**## Procedure and Setting**

All study procedures were carried out in accordance with the ethical standards of the University’s Institutional Review Board and the authors received a positive vote on the study procedures from the Ethics Committee. All participants were informed in detail about the aim and intention of the study prior to testing. Participation was voluntary and only took place after written consent had been given.

Ein Bild, das Text, Diagramm, Reihe, Screenshot enthält.

Automatisch generierte Beschreibung**Fig. 1.**Procedure of the two-hours 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). Within this timeframe, teachers were welcomed to the lab, taught a 15-minute self-prepared unit, filled out questionnaires, and were interviewed about the previously taught unit while watching the video of it. In detail, the five phases were designed as follows: In the (1) pre-teaching phase, the experimenter welcomed the participants and put on the fitness tracker, 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 standing. During the (2) teaching phase, the participants held their self-prepared micro teaching unit to a class of three trained actors, who executed nine typical 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 an event every one and a half minutes. To avoid sequency effects of the order of the events and the performing actors, we used a fully balanced Latin square design. The teaching unit was video-recorded and lasted about 15-20 minutes. Participants spent this time mostly standing. Thereafter followed the (3) post-teaching phase in which the participants provided demographic information and answered questionnaires irrelevant to this study. This phase took approximately 10 minutes and participants were in a seated position. In the (4) interview phase, the participants watched the video of the unit and answered - among others - questions about their cognitive appraisal of the events (see variables 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.

Out of the above-described five phases, we selected five corresponding intervals with a length of 10 minutes each. We determined the length of the intervals to be 10 minutes, because of theoretical and practical reasons. Foremost, previous research indicated that 10-minute intervals are a useful duration for analyzing PPG data [@lu2008can] and 10 minutes was the shortest duration of a phase in this study. Moreover, teachers' HR differs in correspondence to teaching phases [@donker2020associations; @junker2021potential]. The first 10-minute interval, the (1) pre-teaching interval, comprised the first 10 minutes after the fitness tracker was put on. The second interval, the (2) teaching interval, began two minutes after the teachers started the teaching unit. This interval was of highest relevance to our study, therefore we explicitly chose an early 10-minutes interval within the teaching phase, as previous studies revealed that the beginning of a lesson is essential regarding teacher-student interaction [@donker2018quantitative; @claessens2017positive].[[1]](#footnote-1) The (3) post-teaching interval began immediately after the end of the teaching unit. Choosing this interval and the (1) pre-teaching interval, we intended to examine if and how much the HR changed before and after the teaching activity. Referring to the (4) interview interval and to ensure all participants were being interviewed during this interval, we selected the mid-10 minutes between the end of the teaching unit and the time point where the fitness tracker was taken off, as the duration of the post-teaching phase varied across the participants. The (5) end interval comprised the last 10 minutes before the fitness tracker was taken off.

**## Variables**

**### Heart Rate**

To measure the teachers’ HR, we used a wrist-based fitness tracker. The model was Fitbit Charge 4. In line with the manufacturer's instructions [@fitbitnd], the device was attached a finger’s width above the participants’ wrist bone. The tracker flashes green LEDs many times per second and uses light-sensitive photodiodes to measure the volume changes in the capillaries to calculate how many times the heart beats per minute. Accordingly, teachers HR was assessed in beats per minute (BPM).

**### Teaching Experience**

The participants’ teaching experience was assessed as a part of sociodemographic data via an online questionnaire using SoSci Survey (Version 3.1.06; Leiner, 2019). Participants were asked to state their work experience in years (excluding the traineeship year).

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

We assessed teachers’ cognitive appraisal of the 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 events on a 11-point rating scale, ranging from 0 (not at all) to 10 (extremely). Then, we averaged the disruption and confidence appraisals across the nine events.

**## Data analysis**

We conducted all analysis with R [@RStudio2020].

Hypothesis 1a referred to the HR trend over the course of the entire study. First, we z-standardized the BPM values so that the resulting values can be interpreted as differences from the overall HR mean in standard deviation units. Then, we displayed and visually compared the unstandardized and standardized HR trend. Additionally, we calculated means, standard deviations, and range of teachers’ unstandardized and standardized HR over the curse of the entire study and separately for the five intervals.

For testing Hypothesis 1b, which examined the HR levels, 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 HR peak, 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.

For testing Hypothesis 1c, 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 random 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)

In Hypotheses 2, we researched the effects of teaching experience and cognitive appraisal of the events during the micro teaching unit on teachers’ HR levels. In a first step, we investigated the effect of solely teaching experience on the participants’ HR for the five intervals by linear regression models (Hypothesis 2a). In the next step, 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.[[3]](#footnote-3) In the last step, we examined the effects of the three predictors in concert (Hypothesis 2d).

**# Results**

**## Heart rate course**

Means, standard deviations, and range of teachers’ unstandardized and standardized HR are shown in Table XX. Fig. XX a. and b. displays the unstandardized mean HR in BPM and the standardized mean HR, respectively. Referring to hypothesis 1a, HR initially increased, peaked, and then decreased. Comparing the unstandardized and standardized HR trends revealed a high similarity of the overall courses.

Table XX

*Means (M), standard deviations (SD) and Range of Teachers’ HR over the course of the entire study and the five intervals (Unstandardized in BPM/z-standardized)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Interval | *M* | *SD* | Min | Max |
| Overall course | 90.09/0.04 | 15.76/0.99 | 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 |

Figure XX

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

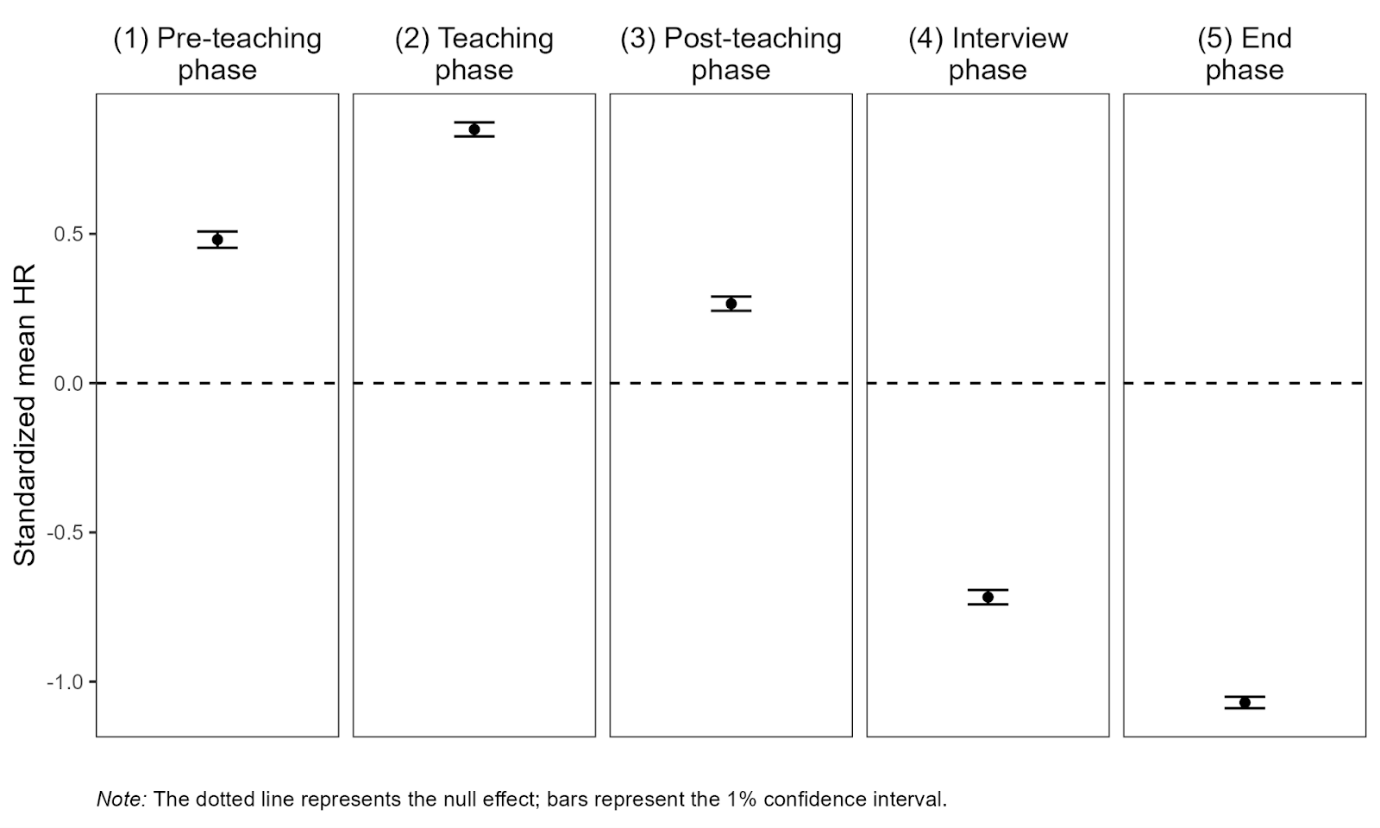


*Note:* The shadow around the line represents the 99% confidence interval. The confidence interval shown refers to the HR measurement points during the entire study period. We used the ggplot2 package (v3.3.3; Wickham, 2016) to calculate the moving average of the course.

Hypothesis 1b aimed to locate the HR peak. Repeated measures ANOVA revealed that the standardized HR means of the intervals differed statistically significant, *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, HR peaked in the (2) teaching interval (see Fig XX).

Figure XX

*Standardized Mean HR for the Five Intervals*

**

Hypothesis 1c, 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 XX; the graphical representation of the slopes is displayed in Figure XX. 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 from zero.

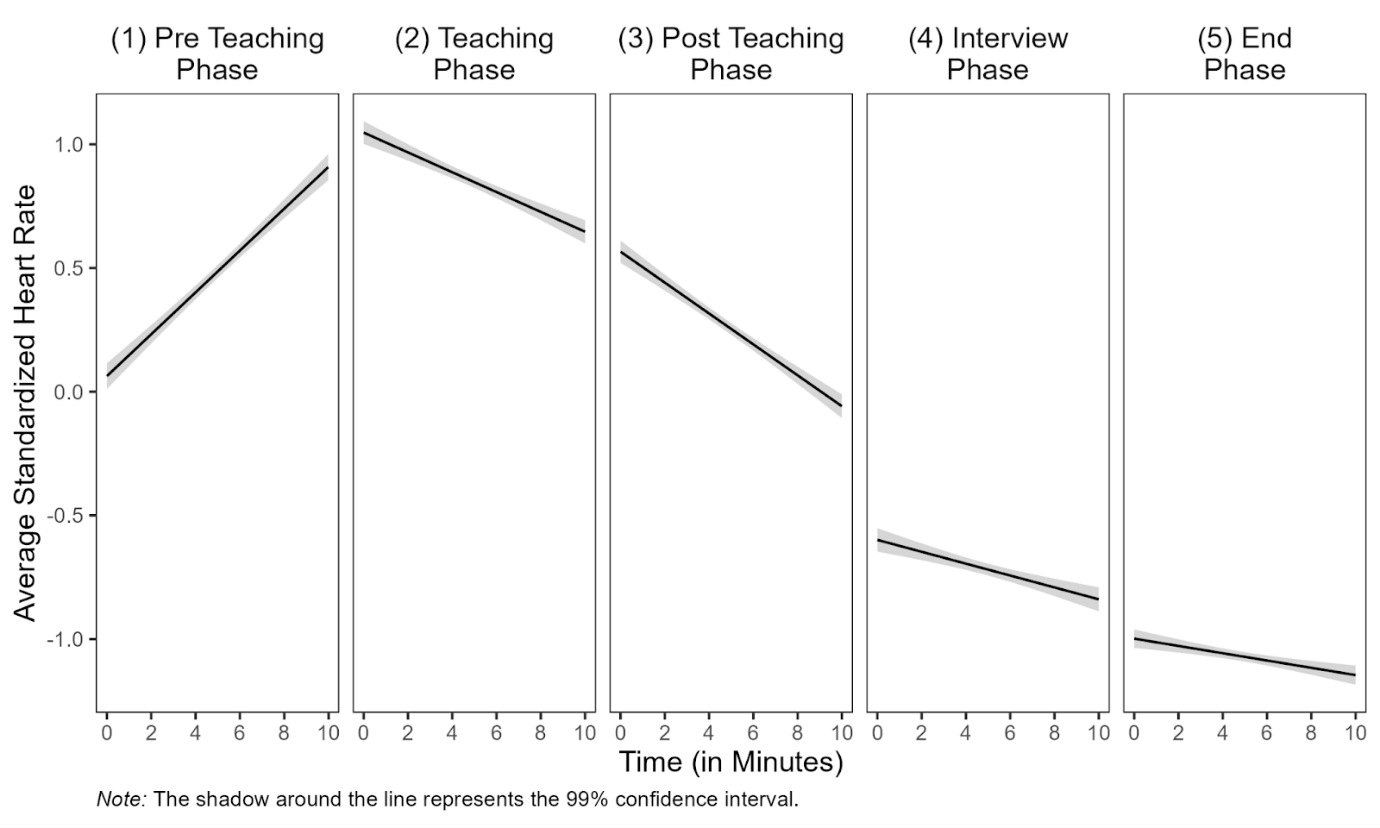
Table XX

*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* stemmed from the variation in the number of collected data points by the fitness tracker. | | | |

Fig. XX

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



**## Prediction of Mean HR 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 XX. Correlations between HR and the other constructs were mostly 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* = 0.012, *p* < .05, Table XX, interview interval). Adding the disruption appraisal while controlling for the shared variance with teaching experience (\*\*Hypothesis 2b\*\*), revealed a significant effect only in the (3) post-teaching interval (*b* = 0.084, *p* < .05, Table XX, post-teaching interval). Adding the confidence appraisal while controlling for the shared variance with teaching experience (\*\*Hypothesis 2c\*\*) showed a significant effect only in the (4) teaching interval (*b* = 0.013, *p* < .05). 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 XX

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | 1 | 2 | 3 | 4 |
| (1) Pre-teaching interval |  |  |  |  |
| 1 HR | − | − .17 | − .02 | − .10 |
| 2 TE |  | − | − .26\* | .34\* |
| 3 DA |  |  | − | − .24\* |
| 4 CA |  |  |  | − |
| (2) Teaching interval |  |  |  |  |
| 1 HR | − | 0.03 | − .19 | − .02 |
| 2 TE |  | − | − .23\* | .36\* |
| 3 DA |  |  | − | − .25\* |
| 4 CA |  |  |  | − |
| (3) Post-teaching interval |  |  |  |  |
| 1 HR | − | .00 | .24\* | .11 |
| 2 TE |  | − | − .23\* | .35\* |
| 3 DA |  |  | − | − .27\* |
| 4 CA |  |  |  | − |
| (4) Interview interval |  |  |  |  |
| 1 HR | − | .21 | − .05 | − .04 |
| 2 TE |  | − | − .22\* | .36\* |
| 3 DA |  |  | − | − .25\* |
| 4 CA |  |  |  | − |
| (5) End interval |  |  |  |  |
| 1 HR | − | .10 | .05 | − .11 |
| 2 TE |  | − | − .25\* | .36\* |
| 3 DA |  |  | − | − .24\* |
| 4 CA |  |  |  | − |
| *Note.* HR = standardized mean heart rate, TE = teaching experience, DA = disruption appraisal, CA = confidence appraisal, \* *p* < .05. | | | | |

Table XX

*Multiple linear regression of Standardized Mean Heart Rate predicted by Teaching Experience, Disruption Appraisal, and Confidence Appraisal for the Five Intervals*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | *Dependent Variable: Standardized Mean HR* | | | | | | | |
|  | Model 1 | | Model 2 | | Model 3 | | Model 4 | |
|  | Est. (SE) | *p* | Est. (SE) | *p* | Est. (SE) | *p* | Est. (SE) | *p* |
| **(1) Pre-teaching Interval** |  |  |  |  |  |  |  |  |
| Teaching  Experience | .001 (.005) | .12 |  |  |  |  |  |  |
| Constant | 0.52\* (0.06) | <.05 |  |  |  |  |  |  |
| **(2) Teaching Interval** |  |  |  |  |  |  |  |  |
| Teaching  Experience | .005 (.005) | .34 | .002 (.005) | .73 | .005  (.006) | .42 | .003  (.006) | .67 |
| Disruption  Appraisal |  |  | -.062 (.041) | .13 |  |  | -.065  (.042) | .13 |
| Confidence  Appraisal |  |  |  |  | .004 (.046) | .92 | -.014 (.047) | .76 |
| Constant | 0.813\* (0.057) | <.05 | 1.150\* (0.227) | <.05 | 0.778\* (0.349) | .03 | 1.274\* (0.471) | .01 |
| **(3) Post-teaching Interval** |  |  |  |  |  |  |  |  |
| Teaching  Experience | .002 (.005) | .70 | .002 (.005) | .76 | -.003 (.006) | .55 | -.001  (.006) | .22 |
| Disruption  Appraisal |  |  | .073 (.040) | .07 |  |  | .084\*  (.041) | .04 |
| Confidence  Appraisal |  |  |  |  | .027 (.045) | .55 | .051 (.046) | .27 |
| Constant | 0.272\* (0.005) | <.05 | -0.122 (0.222) | .59 | 0.069 (0.343) | .84 | -0.570 (0.457) | .22 |
| **(4) Interview Interval** |  |  |  |  |  |  |  |  |
| Teaching  Experience | .012\* (.006) | .03 | .011 (.006) | .06 | .013\* (.006) | .04 | .012  (.007) | .07 |
| Disruption  Appraisal |  |  | -.020 (.044) | .60 |  |  | -.024  (.047) | .61 |
| Confidence  Appraisal |  |  |  |  | -.010 (.050) | .85 | -.016 (.052) | .76 |
| Constant | 0.793\* (0.062) | <.05 | -0.684 (0.252) | <.05 | -0.721 (0.382) | .06 | -0.541 (0.522) | .30 |
| **(5) End Interval** |  |  |  |  |  |  |  |  |
| Teaching  Experience | .002 (.004) | .67 | .003 (.005) | .58 | .004 (.005) | .46 | .004 (.005) | .43 |
| Disruption  Appraisal |  |  | 0.019 (0.035) | .60 |  |  | .011 (.037) | .76 |
| Confidence  Appraisal |  |  |  |  | -.035 (.039) | .38 | -.032 (.041) | .44 |
| Constant | -1.075\* (0.049) | <.05 | -1.176\* (0.199) | <.05 | -0.811 (0.300) |  | -0.897\* (0.411) | .03 |
| *Note*. 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**

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

Automatisch generierte Beschreibung

**Fig. 2.**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, Zeichnung, Entwurf, Tisch enthält.

Automatisch generierte Beschreibung**Fig. 3.** Setting of the interview. *Note.* The experimenter and participant watched the previously taught unit on video.











1. To record this interval, it was necessary for the experimenter to note the time displayed on the fitness tracker. To ensure that the participants' HR was recorded while they had already started teaching, another two minutes were added to the noted time. [↑](#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)
3. In order to avoid a post-diction, the mean HR was 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-3)