

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

3 Mandy Klatt¹, Peer Keßler¹, Gregor Kachel^{1,2}, Christin Lotz¹, & Anne Deiglmayr¹

4 ¹ Leipzig University

5 ² Max Planck Institute for Evolutionary Anthropology

6 Author Note

7 We received funding from QualiFond of University Leipzig. We have no conflicts of
8 interest to disclose. This article is based on data used at conference presentations
9 (DACH-Nachwuchsakademie, 2022; EARLI SIG 11, 2022; EARLI SIG 27, 2022).

10 The authors made the following contributions. Mandy Klatt: Conceptualization,
11 Writing - Original Draft Preparation, Writing - Review & Editing; Peer Keßler: Writing -
12 Original Draft Preparation, Writing - Review & Editing; Gregor Kachel:
13 Conceptualization, Writing - Original Draft Preparation, Writing - Review & Editing;
14 Christin Lotz: Conceptualization, Writing - Original Draft Preparation, Writing - Review
15 & Editing; Anne Deiglmayr: Supervision.

16 Correspondence concerning this article should be addressed to Mandy Klatt,
17 Dittrichring 5-7, 04109 Leipzig. E-mail: mandy.klatt@uni-leipzig.de

Abstract

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

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

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

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

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

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

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

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

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

Word count: X

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

Introduction

Physiological data such as heart rate are becoming increasingly important in research on stress experience. They represent an important indicator of physical or emotional stress, as increased workload is associated with increased heart rate (Sachs, 2014). Furthermore, they allow a more objective recording of stress than self-reports (Runge, Haarman, & Fisher, n.d.). However, capturing heart rate in an educational context requires the use of low-cost and non-invasive instruments. Fitness trackers worn on the wrist have the potential to be such a useful tool (Ferguson, Rowlands, Olds, & Maher, 2015).

To date, there is still little evidence on the usefulness of heart rate measurements using fitness trackers in teaching and learning settings (Ertzberger & Martin, 2016; Lowe, 2016). Runge et al. (n.d.) alone examined teacher stress in a relatively small sample ($N = 4$ teachers) and showed that high heart rate indicates more stress in teachers.

Thus, there remains a lack of robust studies on whether fitness trackers are an efficient, low-cost, and robust measurement method for assessing teachers' experience of arousal during teaching.

Theoretical Background

Stress in Teaching Profession

→ teacher profession is one of the most stressful professions.

Teacher stress can be defined as “[...] the experience by a teacher of unpleasant, negative emotions, such as anger, anxiety, tension, frustration or depression, resulting from some aspect of their work as a teacher.” (Kyriacou, 2001).

Teachers' individual perceptions of student misbehavior in the classroom are closely related to their well-being (Aldrup, Klusmann, Lüdtke, Göllner, & Trautwein, 2018).

→ wie entsteht Stress

→ wie wurde Stress bisher gemessen

Heart rate as an indicator for stress or arousal

Heart rate is physiologically regulated by the autonomic nervous system. An increase in the activity of the sympathetic as part of the autonomic nervous system results in the heart rate being speeded up ("fight or flight"). On the other hand, an increased activity of the parasympathetic as the counterpart has the effect of slowing down the heart rate ("rest and digest") (Battipaglia & Lanza, 2015). In addition to the autonomic nervous system and genetic factors, heart rate is influenced by numerous external factors such as social, personal, psychological, environmental and behavioural factors (Wang, Lizardo, & Hachen, 2022).

Wrist-worn devices as a new approach to assess physiological measures

Fuller et al. (2020) showed in their review article that wearable devices such as Fitbit watches are accurate and reliable for measuring heart rate in controlled settings.

"The use of physiological measures enabled us to get some insight into teachers' affective responses without disrupting the teaching process (Mauss & Robinson, 2009) and to reduce issues with social desirability, retrospective bias, and high cognitive load (Becker et al., 2015; Goetz et al., 2015; Scollon et al., 2009; Wilhelm & Grossman, 2010). Moreover, we found that heart rate measures discriminated between both teachers, even when their interpersonal behavior during the lesson start was relatively similar."

(20) (PDF) A Quantitative Exploration of Two Teachers with Contrasting Emotions:

Intra-Individual Process Analyses of Physiology and Interpersonal Behavior.

Available from: https://www.researchgate.net/publication/329787434_A_Quantitative_Exploration_of_Two_Teachers_with_Contrasting_Emotions_Intra-Individual_Process_Analyses_of_Physiology_and_Interpersonal_Behavior [accessed Dec 07 2022].

Aim of the study

In the present study, we assessed HR measures and self-report data of pre- and in-service teachers in a controlled teaching-learning setting. The aim was to investigate whether heart rate measurements using wrist-worn fitness trackers are a suitable and effective method **(1)** to map differences in states of arousal between five different phases (pre-teaching phase, teaching phase, post-teaching phase, interview phase and end phase) and **(2)** to evaluate the correlation between self-reported evaluations and HR measures.

(H1) We expected heart rates to be higher during the teaching phase than during the pre- and the three post-teaching phases, and that the HR measures would decrease over the course of the study. **(H2)** We also predicted that HR and a high ranking on the negative scale on our survey (feeling disturbed by disruptions) would correlate positively and a high ranking on the positive scale (feeling confident in dealing with disruptions) would follow the inverse pattern.

Methods

We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study.

Participants

The sample consisted in total of $N = 76$ pre- and in-service teachers. The subjects were recruited from the Leipzig University or from German schools in Saxony via personal

Table 1

Demographic Information

Variable	values
N	76.00
Gender female in percent	63.16
M Age in years	30.26
SD Age in years	10.52
Min Age in years	19.00
Max Age in years	59.00
M Teaching Experience in years	5.17
SD Teaching Experience in years	9.12
Min Teaching Experience in years	0.00
Max Teaching Experience in years	37.00

contact, e-mail lists and flyers. Data of two participants were excluded from the analyses due to insufficient data quality, yielding data from $N = 76$ subjects.

The subjects (48 women; 63.16 %) had a mean age of 30.30 years ($SD = 10.50$; range: 19-59) and an average teaching experience of 5.17 years ($SD = 9.12$; range: 0-37).

15.79% of the subjects were (studying to become) teachers for primary school, 71.05% were (studying to become) teachers for secondary school and 10.53% were (studying to become) teachers for special education needs.

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 Board of Leipzig University. All participants were informed in detail about the aim and intention of the study prior to testing. Participation in the study was voluntary and only took place after written consent has been given.

Material

Teachers' heart rate. We used a Fitbit Charge 4 to measure the teachers' heart rate. The device was attached 2-finger widths above the ulnar styloid process to the

subject's wrist following the manufacturer's instructions. To determine the heart rate, the Fitbit flashes green LEDs many times per second and uses light-sensitive photodiodes to measure the volume changes in the capillaries and then calculates how many times the heart beats per minute (bpm). Data were automatically wireless synced with an iPad via Bluetooth to a Fitbit account, and subsequently, the intraday second-by-second data were exported for each session using the opensource software Pulse Watch (PulseWatch. URL: <https://iccir919.github.io/pulseWatch/public/index.html> [accessed 2022-08-03]).

Teachers' self-reported data of arousal during the teaching phase. The subject's self-reported data of arousal in response to the nine disruptions during the lesson was assessed in a Stimulated Recall Interview that took place after the lesson. For this purpose, two numerical 11-point rating scales were used: (1) The first rating scale collected data of the teacher's subjective perception of disruptiveness of each disruption ("On a scale of 0-10: how disruptive did you find the event? 0 is not disruptive at all, 10 is extremely disruptive") (2) The second rating scale assessed the teacher's subjective perception of confidence in dealing with the disruptions during the lesson ("How confident did you feel in dealing with this event on a scale of 0-10? 0 is not confident at all and 10 is extremely confident").

The response format was purposely chosen to be closed and with several answer options in order to assign numerical values to the subjects' self-report and emotional experience, and thus to make them measurable and comparable (Döring & Bortz, 2016; Eid, Gollwitzer, & Schmitt, 2015). The gradations of the rating scale were unambiguous and the intervals between them identical. The characteristic value was estimated by the subjects immediately after seeing the recording of the respective disruption and communicated verbally to the experimenter.

Procedure

The data collection was part of a larger research project with a planned sample of $n = 40$ in-service teachers and $n = 40$ pre-service teachers.

The study took place in the rooms of the Faculty of Education at Leipzig University. In a controlled laboratory setting, heart rate data in beats per minute (bpm) were recorded using Fitbit Charge 4 over a total period of approximately two hours. Within this time frame, teachers taught a 15-minute self-prepared lesson to an audience of three actors.

For this purpose, a seminar room was converted into a classroom. The classroom was equipped with a digital whiteboard and a blackboard. The fictitious class consisting of three trained actors sat facing the teacher head-on in a “U”; pen and paper were provided on their seats. In addition, the students brought their own mobile phone, which was also visibly positioned on the desk.

The subject was instructed during the lesson to behave and move as naturally as possible, as they would in a real classroom. In advance, all subjects were given information about planning their lesson in a meeting to ensure that an appropriately realistic teaching situation could take place, e.g. it was pointed out that longer film clips as well as group work as a social form should be avoided in order to ensure interaction between teacher and students in the short time of the study.

For analyzing the heart rate data, we focused on five 10-minute intervals of theoretical interest for multiple reason. First, 10 minutes was to the minimum duration of all intervals, so we wanted to ensure the comparability of the intervals for all participants. Second, Lu et al. (2008) confirmed in their study that 10-minute intervals are an useful duration for analyzing photoplethysmography (PPG) data. Third, other studies revealed that the first minutes of the lesson start are essential regarding teacher-student interaction (Claessens et al., 2017; Donker, Van Gog, & Mainhard, 2018).

The procedure of the study and how we calculated the five phases is described in the

following:

(1) The *pre-teaching phase* was the first 10-minute interval of interest. This interval was calculated from the moment the subject arrived and the Fitbit watch was put on, which happened immediately after the subject was welcomed by the experimenter and the three actors. After the experimenter briefly explained the project and the procedure of the study to the subject, contact details were subsequently collected as part of the hygiene concept of COVID-19 and the written consent to voluntarily participate in the study was requested. Next, the subject was asked to prepare the necessary materials for the lesson (connecting the laptop to the beamer, preparing worksheets, etc.). This phase varied in duration according to the prepared lesson (about 1-5min on average).

Once the preparation was completed, the eye-tracking glasses were briefly explained to the subject as well as the adjustment made to the subject by the experimenter (changing the lenses or the nose pad if necessary). Subsequently, a 1-point calibration of the eye-tracking glasses took place and together with further devices (GoPro camera, 3D camera, audio recorder) the recordings were started.

Before the lesson began, a warm-up phase took place to familiarize the subject with the laboratory setting and equipment. This warm-up phase consisted of two parts:

- In the first part, the subject and the three actors playfully learned each other's names. Additional fictitious name tags of the actors were attached to the table and did not change across the study. In the game, by throwing two balls, either the participant's own name or the name of the target person was called.
- The second part of the warm-up phase involved getting into conversation with each other by asking questions. For this purpose, the subject was asked to come up with a question for each of the actors (three questions in total) and was also asked a question in return by each actor. This could be anything that interested the participants. The questions should be as authentic as possible and were not tailored

to the role of the actors. The purpose of this warming-up phase was, on the one hand, to make the subject forget about the devices and, on the other hand, to create as realistic as possible a classroom situation in which interactions were based on a teacher-student relationship.

The pre-teaching phase lasted 10-15 minutes on average, depending on different issues (preparing the lesson, technical problems, adjusting the eye-tracking glasses).

(2) The second 10-minute interval was the *teaching phase*, which began with the experimenter noting the time and step count of the fitness tracker. The subject was then asked to carry out a 9-point calibration alone in a room next door. After the calibration, the lesson started with the subject opening the door to the laboratory room. During the lesson, the actors were instructed to simulate nine typical classroom events, three each in the following categories: (a) verbal disruptive behavior (chatting with the neighbor, whispering, heckling), (b) physical disruptions (clicking with a pen, drumming with hands on the table, snipping) and (c) lack of eagerness to learn (drawing on a sheet of paper, putting the head on the table, looking at the phone).

The order of the events and the performing actors were fully balanced with Latin square. The actors were trained before to perform the disruptions identically in each data collection.

After a short familiarization phase for the teacher of two and a half minutes, the instructions appeared as intervals (every 90 seconds for 30 seconds) on a screen that was only visible for the class. The actors were trained to stop the disruptive behavior as soon as the teacher intervened.

Time management was regulated by the experimenter by showing time cards (yellow = 1 minute until the end; red = end of the lesson). The lesson lasted about 18 minutes on average but only the first 10 minutes were used as an interval to ensure the identical duration for all participants. At the end of the lesson, the experimenter again noted the

time and the steps on the clock.

(3) The noted time of the lesson end was also the starting point for the *post-teaching* phase as the third 10-minute interval. At the end of the lesson, the experimenter attached letters A through D in the lab while the teacher performed a second 9-point calibration outside. The letter search as a control condition was part of another project, which is why it will not be discussed further in the study. Subsequently, the subjects as well as the actors were given a short questionnaire, which contained items to collect demographic information as well as items about the previously given lesson on teaching quality (4-point Likert scale; 1 = strongly disagree; 2 = disagree; 3 = agree; 4 = strongly agree; EMU, Helmke et al. (2014)) via the online survey website SoSci Survey. The completion of the questionnaire took approximately 5 minutes.

Meanwhile, the experimenter prepared the second part of the data collection: The Stimulated Recall Interview. Within post-teaching phase, the majority of subjects completed the questionnaire and positioned themselves next to the experimenter to begin the interview. In some cases, the beginning of the interview was part of the post-teaching phase due to individual needs (going to the toilet, spending more time completing the questionnaire, etc.).

(4) The fourth 10-minute interval of interest was the *interview phase* that was 10 minutes in the middle of the interview. To ensure that all subjects were in the middle of the interview, we calculated the difference from the end of the lesson and from the time when the subject took off the watch. This duration was divided in two to get to the middle of the interval. Then, 5 minutes were subtracted to get to the start of the 10-minute interval.

The interview itself took place using the Think Aloud procedure (Konrad, 2020). For this purpose, the eye tracking video was watched again and the experimenter stopped the video at the nine disruptions when they appeared to ask several questions. The questions were identical for each disruption, they were asked in a fixed order. First, the subject was

asked to describe the disruption, to evaluate (11-point rating scale) and justify the disruptiveness of each disruption. Next, the subject was asked to describe and justify the reaction. The experimenter then asked the subject to evaluate (11-point rating scale) and justify the confidence the subject had in dealing with the disruption. Statements about the evaluation of the disruptiveness and the confidence were quantified. For this purpose, subjects determined their individual value on a rating scale from 0 (not disturbing / not safe) to 10 (disturbing / safe).

The interview lasted on average 45-60 minutes.

(5) The last interesting interval for the analyses was the *end phase*, which was the 10 minutes before the subject took off the watch. In this interval, the subject answered a second questionnaire (Situational Judgment Test, Gold and Holodyski (2015)) to assess the subject's strategic knowledge of classroom management. The subject was asked to judge alternative actions on school scenarios on a 6-point rating scale from 1 (A) to 6 (F) according to school grades. Data from the questionnaire were again collected as an online questionnaire via the website www.soscisurvey.de and lasted approximately 10-15 minutes.

The Fitbit watch was removed only after the last questionnaire to obtain heart rate data during the entire study.

Data analysis

We used R (Version 4.2.2; R Core Team, 2022) and the R-packages *broom* (Version 1.0.1; Robinson, Hayes, & Couch, 2022), *cowplot* (Version 1.1.1; Wilke, 2020), *DescTools* (Version 0.99.48; Andri et mult. al., 2022), *dplyr* (Version 1.0.10; Wickham, François, Henry, & Müller, 2022), *forcats* (Version 0.5.2; Wickham, 2021), *ggplot2* (Version 3.3.6; Wickham, 2016), *ggpubr* (Version 0.4.0; Kassambara, 2020), *ggthemes* (Version 4.2.4; Arnold, 2021), *gridExtra* (Version 2.3; Auguie, 2017), *imputeTS* (Version 3.3; Moritz & Bartz-Beielstein, 2017), *janitor* (Version 2.2.0; Firke, 2021), *jtools* (Version 2.2.1; Long,

2022), *lm.beta* (Version 1.7.2; Behrendt, 2022), *lme4* (Version 1.1.31; Bates, Mächler,
 Bolker, & Walker, 2015), *ltm* (Version 1.2.0; Rizopoulos, 2006), *lubridate* (Version 1.8.0;
 Grolemund & Wickham, 2011), *MASS* (Version 7.3.58.1; Venables & Ripley, 2002), *Matrix*
 (Version 1.5.1; Bates, Maechler, & Jagan, 2022), *msm* (Version 1.7; Jackson, 2011), *needs*
 (Version 0.0.3; Katz, 2016), *papaja* (Version 0.1.1.9001; Aust & Barth, 2020), *polycor*
 (Version 0.8.1; Fox, 2022), *ppcor* (Version 1.1; Kim, 2015), *psych* (Version 2.3.3; Revelle,
 2022), *purrr* (Version 0.3.5; Henry & Wickham, 2020), *readr* (Version 2.1.3; Wickham,
 Hester, & Bryan, 2022), *readxl* (Version 1.4.1; Wickham & Bryan, 2022), *rstatix* (Version
 0.7.1; Kassambara, 2021), *rticles* (Version 0.25; Allaire et al., 2023), *stringr* (Version 1.4.1;
 Wickham, 2019), *tibble* (Version 3.1.8; Müller & Wickham, 2021), *tidyr* (Version 1.2.1;
 Wickham & Girlich, 2022), *tidyverse* (Version 1.3.2; Wickham et al., 2019), *tinylabels*
 (Version 0.2.3; Barth, 2022), *viridis* (Version 0.6.2; Garnier et al., 2021a, 2021b), *viridisLite*
 (Version 0.4.1; Garnier et al., 2021b), and *xfun* (Version 0.38; Xie, 2022) for all our
 analyses (descriptive statistics, correlations, and multilevel regression). Statistical
 assumptions including linearity, homogeneity of variance, and normal distribution of
 residuals were tested and confirmed with RStudio [palmeri2020]. An alpha level of 0.05 was
 used for the statistical tests.

Descriptive statistics. Means, standard deviations and the range of teachers'
 heart rate and the self-reported data are shown in Table 2.

Teachers' average heart rate during the entire course of the study was approximately
 90.40 bpm with a standard deviation of 15.82 and a range from 51 to 164 bpm.

Means, standard deviations and the range of teachers' heart rate for the different
 phases and the self-reported data are shown in Table 2.

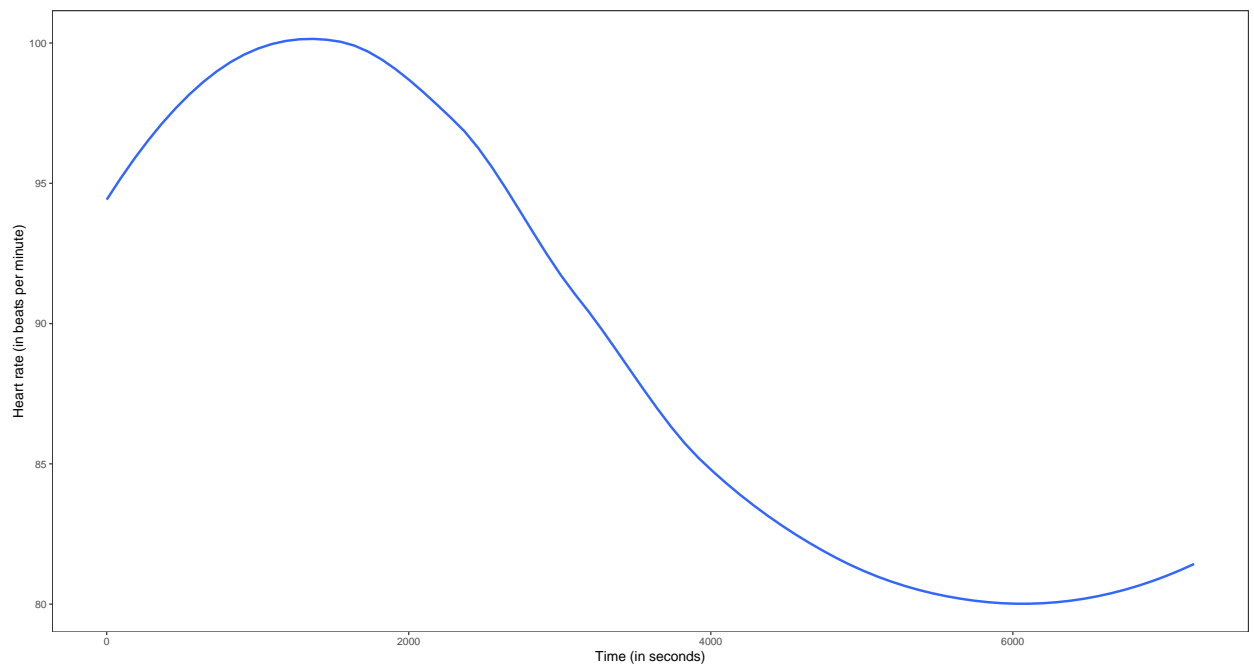
Pairwise t-test and effect size.

Multilevel regressions. XXX

Results

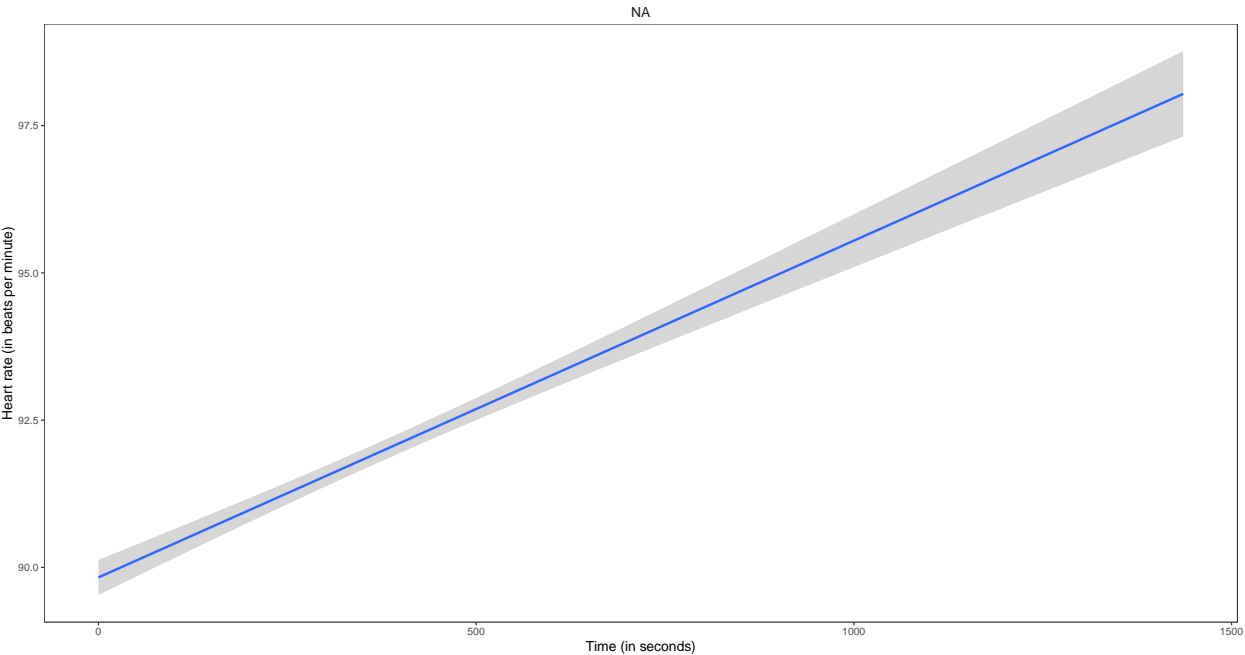
First visual results

The graph @ref(fig:overall_plot) show the heart rate for the entire course of the study.



The second graph @ref(fig:phase_plot) show the heart rate for the individual five phases.

Warning: Unknown levels in 'f': pre, main, subsequent, interview, after



310

311

Discussion and implications

References

- Aldrup, K., Klusmann, U., Lüdtke, O., Göllner, R., & Trautwein, U. (2018). Student misbehavior and teacher well-being: Testing the mediating role of the teacher-student relationship. *Learning and Instruction, 58*, 126–136.
<https://doi.org/https://doi.org/10.1016/j.learninstruc.2018.05.006>
- Allaire, J., Xie, Y., Dervieux, C., R Foundation, Wickham, H., Journal of Statistical Software, ... Hyndman, R. (2023). *Rticles: Article formats for r markdown*. Retrieved from <https://CRAN.R-project.org/package=rticles>
- Andri et mult. al., S. (2022). *DescTools: Tools for descriptive statistics*. Retrieved from <https://cran.r-project.org/package=DescTools>
- Arnold, J. B. (2021). *Ggthemes: Extra themes, scales and geoms for 'ggplot2'*. Retrieved from <https://CRAN.R-project.org/package=ggthemes>
- Auguie, B. (2017). *gridExtra: Miscellaneous functions for "grid" graphics*. Retrieved from <https://CRAN.R-project.org/package=gridExtra>
- Aust, F., & Barth, M. (2020). *papaja: Prepare reproducible APA journal articles with R Markdown*. Retrieved from <https://github.com/crsh/papaja>
- Barth, M. (2022). *tinylabels: Lightweight variable labels*. Retrieved from <https://cran.r-project.org/package=tinylabels>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.
<https://doi.org/10.18637/jss.v067.i01>
- Bates, D., Maechler, M., & Jagan, M. (2022). *Matrix: Sparse and dense matrix classes and methods*. Retrieved from <https://CRAN.R-project.org/package=Matrix>
- Battipaglia, I., & Lanza, G. A. (2015). The autonomic nervous system of the heart. In R. H. J. A. Slart, R. A. Tio, P. H. Elsinga, & M. Schwaiger (Eds.), *Autonomic innervation of the heart: Role of molecular imaging* (pp. 1–12). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-45074-1_1

- Behrendt, S. (2022). *Lm.beta: Add standardized regression coefficients to lm-objects*. Retrieved from <https://CRAN.R-project.org/package=lm.beta>
- Claessens, L. C., Tartwijk, J. van, Want, A. C. van der, Pennings, H. J., Verloop, N., Brok, P. J. den, & Wubbels, T. (2017). Positive teacher–student relationships go beyond the classroom, problematic ones stay inside. *The Journal of Educational Research*, 110(5), 478–493.
- Donker, M. H., Van Gog, T., & Mainhard, M. T. (2018). A quantitative exploration of two teachers with contrasting emotions: Intra-individual process analyses of physiology and interpersonal behavior. *Frontline Learning Research*, 6(3), 162–184.
- Döring, N., & Bortz, J. (2016). Empirische sozialforschung im Überblick. In *Forschungsmethoden und evaluation in den sozial-und humanwissenschaften* (pp. 3–30). Springer.
- Eid, M., Gollwitzer, M., & Schmitt, M. (2015). *Statistik und forschungsmethoden: Mit online-materialien* (4., überarb. Und erw. aufl.). Weinheim: Beltz. Retrieved from https://contentselect.com/media/moz_viewer
- Ertzberger, J., & Martin, F. (2016). Use of fitness bands by teachers in the classroom. *TechTrends*, 60(4), 392–397.
- Ferguson, T., Rowlands, A. V., Olds, T., & Maher, C. (2015). The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: A cross-sectional study. *International Journal of Behavioral Nutrition and Physical Activity*, 12(1), 1–9.
- Firke, S. (2021). *Janitor: Simple tools for examining and cleaning dirty data*. Retrieved from <https://CRAN.R-project.org/package=janitor>
- Fox, J. (2022). *Polycor: Polychoric and polyserial correlations*. Retrieved from <https://CRAN.R-project.org/package=polycor>
- Fuller, D., Colwell, E., Low, J., Orychock, K., Tobin, M. A., Simango, B., et al.others. (2020). Reliability and validity of commercially available wearable devices for

measuring steps, energy expenditure, and heart rate: Systematic review. *JMIR mHealth and uHealth*, 8(9), e18694.

Garnier, Simon, Ross, Noam, Rudis, Robert, . . . Cédric. (2021a). *viridis - colorblind-friendly color maps for r*. <https://doi.org/10.5281/zenodo.4679424>

Garnier, Simon, Ross, Noam, Rudis, Robert, . . . Cédric. (2021b). *viridis - colorblind-friendly color maps for r*. <https://doi.org/10.5281/zenodo.4679424>

Gold, B., & Holodynski, M. (2015). Development and construct validation of a situational judgment test of strategic knowledge of classroom management in elementary schools. *Educational Assessment*, 20(3), 226–248.

Grolemund, G., & Wickham, H. (2011). Dates and times made easy with lubridate. *Journal of Statistical Software*, 40(3), 1–25. Retrieved from <https://www.jstatsoft.org/v40/i03/>

Helmke, A., Helmke, T., Lenske, G., Pham, G., Praetorius, A.-K., Schrader, F.-W., & AdeThuwrow, M. (2014). Unterrichtsdiagnostik mit EMU. *Aus-Und Fortbildung Der Lehrkräfte in Hinblick Auf Verbesserung Der Diagnosefähigkeit, Umgang Mit Heterogenität Und Individuelle Förderung*, 149–163.

Henry, L., & Wickham, H. (2020). *Purrr: Functional programming tools*. Retrieved from <https://CRAN.R-project.org/package=purrr>

Jackson, C. H. (2011). Multi-state models for panel data: The msm package for R. *Journal of Statistical Software*, 38(8), 1–29. Retrieved from <https://www.jstatsoft.org/v38/i08/>

Kassambara, A. (2020). *Ggpubr: 'ggplot2' based publication ready plots*. Retrieved from <https://CRAN.R-project.org/package=ggpubr>

Kassambara, A. (2021). *Rstatix: Pipe-friendly framework for basic statistical tests*. Retrieved from <https://CRAN.R-project.org/package=rstatix>

Katz, J. (2016). *Needs: Attaches and installs packages*. Retrieved from <https://CRAN.R-project.org/package=needs>

Kim, S. (2015). *Ppcor: Partial and semi-partial (part) correlation*. Retrieved from

393 <https://CRAN.R-project.org/package=ppcor>

394 Kyriacou, C. (2001). Teacher stress: Directions for future research. *Educational Review*,
395 53(1), 27–35.

396 Long, J. A. (2022). *Jtools: Analysis and presentation of social scientific data*. Retrieved
397 from <https://cran.r-project.org/package=jtools>

398 Lowe, H. (2016). Educational behavior apps and wearable devices: Current research and
399 prospects. *Educational Technology*, 30–33.

400 Lu, S., Zhao, H., Ju, K., Shin, K., Lee, M., Shelley, K., & Chon, K. H. (2008). Can
401 photoplethysmography variability serve as an alternative approach to obtain heart rate
402 variability information? *Journal of Clinical Monitoring and Computing*, 22(1), 23–29.

403 Moritz, S., & Bartz-Beielstein, T. (2017). imputeTS: Time Series Missing Value
404 Imputation in R. *The R Journal*, 9(1), 207–218. <https://doi.org/10.32614/RJ-2017-009>

405 Müller, K., & Wickham, H. (2021). *Tibble: Simple data frames*. Retrieved from
406 <https://CRAN.R-project.org/package=tibble>

407 R Core Team. (2022). *R: A language and environment for statistical computing*. Vienna,
408 Austria: R Foundation for Statistical Computing. Retrieved from
409 <https://www.R-project.org/>

410 Revelle, W. (2022). *Psych: Procedures for psychological, psychometric, and personality*
411 *research*. Evanston, Illinois: Northwestern University. Retrieved from
412 <https://CRAN.R-project.org/package=psych>

413 Rizopoulos, D. (2006). Ltm: An r package for latent variable modelling and item response
414 theory analyses. *Journal of Statistical Software*, 17(5), 1–25. Retrieved from
415 <https://doi.org/10.18637/jss.v017.i05>

416 Robinson, D., Hayes, A., & Couch, S. (2022). *Broom: Convert statistical objects into tidy*
417 *tibbles*. Retrieved from <https://CRAN.R-project.org/package=broom>

418 Runge, N., Haarman, S., & Fisher, M. H. (n.d.). *Using fitbit fitness trackers to measure*
419 *teacher stress and coping*.

- Sachs, S. (2014). *Physiologische parameter zur bewertung der lernwirksamkeit von lernsituationen* (PhD thesis). Universität Ulm.
- Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with s* (Fourth). New York: Springer. Retrieved from <https://www.stats.ox.ac.uk/pub/MASS4/>
- Wang, C., Lizardo, O., & Hachen, D. S. (2022). Using fitbit data to monitor the heart rate evolution patterns of college students. *Journal of American College Health*, 70(3), 875–882.
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from <https://ggplot2.tidyverse.org>
- Wickham, H. (2019). *Stringr: Simple, consistent wrappers for common string operations*. Retrieved from <https://CRAN.R-project.org/package=stringr>
- Wickham, H. (2021). *Forcats: Tools for working with categorical variables (factors)*. Retrieved from <https://CRAN.R-project.org/package=forcats>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., . . . Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- Wickham, H., & Bryan, J. (2022). *Readxl: Read excel files*. Retrieved from <https://CRAN.R-project.org/package=readxl>
- Wickham, H., François, R., Henry, L., & Müller, K. (2022). *Dplyr: A grammar of data manipulation*. Retrieved from <https://CRAN.R-project.org/package=dplyr>
- Wickham, H., & Girlich, M. (2022). *Tidyr: Tidy messy data*. Retrieved from <https://CRAN.R-project.org/package=tidyr>
- Wickham, H., Hester, J., & Bryan, J. (2022). *Readr: Read rectangular text data*. Retrieved from <https://CRAN.R-project.org/package=readr>
- Wilke, C. O. (2020). *Cowplot: Streamlined plot theme and plot annotations for 'ggplot2'*. Retrieved from <https://CRAN.R-project.org/package=cowplot>
- Xie, Y. (2022). *Xfun: Supporting functions for packages maintained by 'yihui xie'*.

447 Retrieved from <https://CRAN.R-project.org/package=xfun>