

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.

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, high step count, and little sleep indicate more stress in teachers.

Thus, there remains a lack of robust studies on whether fitness trackers are an efficient, inexpensive, and robust measurement method for assessing teachers' experience of stress. Since, among other things, greater professional experience is associated with lower feelings of stress in the teaching profession (Fisher, 2011), this study will also consider the heart rate measurements of experienced and inexperienced teachers.

Aim of the study

Against this background, we investigated whether heart rate measurements using fitness trackers in a controlled teaching-learning setting are suitable to map differences between (1) anticipation, arousal, and rest phases. We expected heart rates to be higher in

the arousal phase than in the rest phases.

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 $N = 63$ pre- and in-service teachers. The subjects were recruited from the Leipzig University or from German schools in Saxony via personal contact, e-mail lists and flyers. Data of two participants were excluded from the analyses due to insufficient data quality, yielding data from $N = 61$ subjects.

The subjects (39 women; 63.93 %) had a mean age of 29.60 years ($SD = 10.40$; range: 19-59) and an average teaching experience of 4.66 years ($SD = 9.30$; range: 0-37).

18.03% of the subjects were (studying to become) teachers for primary school, 72.13% were (studying to become) teachers for secondary school and 9.84% 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-assessment of arousal during the teaching phase. The subject's self-assessment 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 with the following question: "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 asked as follows: "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-assessment 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 selected five 10-minute intervals of theoretical interest. To ensure that the intervals are comparable for all participants, we decided on 10-minute intervals because this was the minimum duration of all intervals. The individual intervals are described in detail below.

(1) The *pre-teaching phase* was the first 10 minutes belonging to the 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). XXX
GENAUEN M + SD BERECHNEN XXX

(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. XXX GENAUEN M + SD BERECHNEN XXX 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 are 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 is 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 4.1.3; 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 0.99.45; Andri et mult. al., 2022), *dplyr* (Version 1.0.10; Wickham, François, Henry, & Müller, 2022), *forcats* (Version 0.5.1; Wickham, 2021), *ggplot2* (Version 3.3.5; Wickham, 2016), *ggpubr* (Version 0.4.0; Kassambara, 2020), *ggthemes* (Version 4.2.4; Arnold, 2021), *gridExtra* (Version 2.3; Auguie, 2017), *imputeTS* (Version 3.2; Moritz & Bartz-Beielstein, 2017), *janitor* (Version 2.1.0; Firke, 2021), *jtools* (Version 2.2.0; Long, 2022), *lm.beta* (Version 1.6.2; Behrendt, 2022), *lme4* (Version 1.1.30; Bates, Mächler, Bolker, & Walker, 2015), *ltm* (Version 1.2.0; Rizopoulos, 2006), *lubridate* (Version 1.8.0; Grolemund & Wickham, 2011), *MASS* (Version 7.3.55; Venables & Ripley, 2002), *Matrix* (Version 1.5.1; Bates, Maechler, & Jagan, 2022), *msm* (Version 1.6.9; Jackson, 2011), *needs* (Version 0.0.3; Katz, 2016), *papaja* (Version 0.1.0.9999; Aust & Barth, 2020), *polycor* (Version 0.8.1; Fox, 2022), *ppcor* (Version 1.1; Kim, 2015), *purrr* (Version 0.3.4; Henry &

Wickham, 2020), *readr* (Version 2.1.2; Wickham, Hester, & Bryan, 2022), *readxl* (Version 1.4.0; Wickham & Bryan, 2022), *rstatix* (Version 0.7.0; Kassambara, 2021), *stringr* (Version 1.4.0; Wickham, 2019), *tibble* (Version 3.1.6; Müller & Wickham, 2021), *tidyr* (Version 1.2.0; Wickham & Girlich, 2022), *tidyverse* (Version 1.3.1; Wickham et al., 2019), *tinylabels* (Version 0.2.3; Barth, 2022), *viridis* (Version 0.6.2; Garnier et al., 2021a, 2021b), and *viridisLite* (Version 0.4.0; Garnier et al., 2021b) for all our analyses.

FitBit. All participants were given a FitBit Smart Watch Charge 4 to wear during the experiment.

Heart Rate

Results

Discussion and implication

References

- 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). *tinylabls: Lightweight variable labels*. Retrieved from <https://cran.r-project.org/package=tinylabls>
- 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>
- Behrendt, S. (2022). *Lm.beta: Add standardized regression coefficients to lm-objects*. Retrieved from <https://CRAN.R-project.org/package=lm.beta>
- 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>

Fisher, M. H. (2011). Factors influencing stress, burnout, and retention of secondary teachers. *Current Issues in Education*, 14(1).

Fox, J. (2022). *Polycor: Polychoric and polyserial correlations*. Retrieved from <https://CRAN.R-project.org/package=polycor>

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., & AdeThuwow, 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 <https://CRAN.R-project.org/package=ppcor>
- Long, J. A. (2022). *Jtools: Analysis and presentation of social scientific data*. Retrieved from <https://cran.r-project.org/package=jtools>
- Lowe, H. (2016). Educational behavior apps and wearable devices: Current research and prospects. *Educational Technology*, 30–33.
- Moritz, S., & Bartz-Beielstein, T. (2017). imputeTS: Time Series Missing Value Imputation in R. *The R Journal*, 9(1), 207–218. <https://doi.org/10.32614/RJ-2017-009>
- Müller, K., & Wickham, H. (2021). *Tibble: Simple data frames*. Retrieved from <https://CRAN.R-project.org/package=tibble>
- R Core Team. (2022). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Rizopoulos, D. (2006). Ltm: An r package for latent variable modelling and item response theory analyses. *Journal of Statistical Software*, 17(5), 1–25. Retrieved from <https://doi.org/10.18637/jss.v017.i05>
- Robinson, D., Hayes, A., & Couch, S. (2022). *Broom: Convert statistical objects into tidy tibbles*. Retrieved from <https://CRAN.R-project.org/package=broom>
- Runge, N., Haarman, S., & Fisher, M. H. (n.d.). *Using fitbit fitness trackers to measure 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/>

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>