

# In-Gauge and En-Gage: Understanding Occupants' Behaviour, Engagement, Emotion, and Comfort Indoors with Heterogeneous Sensors and Wearables

Nan Gao , Max Marschall , Jane Burry , Simon Watkins , Flora Salim 

Published: Feb. 13, 2023. Version: 1.0.0

**When using this resource, please cite:** [\(show more options\)](#)

Gao, N., Marschall, M., Burry, J., Watkins, S., & Salim, F. (2023). In-Gauge and En-Gage: Understanding Occupants' Behaviour, Engagement, Emotion, and Comfort Indoors with Heterogeneous Sensors and Wearables (version 1.0.0). *PhysioNet*. <https://doi.org/10.13026/srm3-7z33>.

**Additionally, please cite the original publication:**

[Gao, N., Marschall, M., Burry, J. et al. Understanding occupants' behaviour, engagement, emotion, and comfort indoors with heterogeneous sensors and wearables. \*Sci Data\* 9, 261 \(2022\). <https://doi.org/10.1038/s41597-022-01347-w>](#)

**Please include the standard citation for PhysioNet:** [\(show more options\)](#)

Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P. C., Mark, R., ... & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation* [Online]. 101 (23), pp. e215–e220.

## Abstract

We conducted a field study at a K-12 private school in the suburbs of Melbourne, Australia. The datasets contained two elements: (1) In-Gauge dataset: we conducted a 5-month longitudinal field study using two outdoor weather stations, as well as **indoor weather stations** in 17 classrooms and **temperature sensors** on the vents of occupant-controlled room air-conditioners; these were collated into individual datasets for each classroom at a 5-minute logging frequency, including additional data on occupant presence. (2) En-Gage dataset: we tracked 23 students and 6 teachers in a 4-week cross-sectional study En-Gage, using wearable sensors to log physiological data (**electrodermal activity, heart rate, blood column pulse, skin temperature, 3-axis acceleration**), as well as daily surveys to query the occupants' **thermal comfort, learning engagement, emotions** and **seating behaviours**. Overall, the combined dataset could be used to analyse the relationships between indoor/outdoor climates and students' behaviours/mental states on campus, which provide opportunities for the future design of intelligent feedback systems to benefit both students and staff.

## Background

How can indoor spaces be designed in ways that increase occupant well-being while decreasing energy consumption? Answering this question requires a holistic understanding of indoor climates, occupant comfort and behaviour, as well as the dynamic relationships between these different aspects. The present study sits within a context of research that aims to gain insights by examining these themes using mixed methods of data capture within operational buildings. More specifically, the study contains two separate assays, each relating to a distinct body of existing research.

The first assay is a 5-month longitudinal field study using outdoor and indoor weather stations as well as sensors to determine the use of occupant-controlled room air-conditioners. This assay was undertaken to contribute knowledge to the research field of occupant behaviour modelling in building performance simulation. During the design of buildings, engineers often use simulations to predict the indoor environmental quality and energy consumption of design options in order to inform decision-making. There are often large discrepancies between simulated and actual building performance [\[1\]](#).

The second assay is a four-week study tracking 23 students and six teachers, using wearable sensors to log physiological data as well as self-report from participants at school. Studying student engagement, emotions, and daily behaviours has attracted increasing interest to address problems such as low academic performance and disaffection. Sensor-based physiological and behaviour recordings provide great opportunities to unobtrusively measure students' behaviours and emotional changes in classroom settings. In previous studies, various physiological signals, such as electrodermal activity (EDA) and heart rate variability (HRV), and environmental data have been

explored to assess emotional arousal [2] and engagement levels [3]. Existing datasets in affective computing either provide limited scope for understanding emotion responses in real-world settings or only consider a particular type of annotation to meet their research goals (e.g., stress level and mental workload).

---

## Methods

For the In-Gauge dataset, the longitudinal study was conducted for a 5.5-month period from the end of 2019 to early 2020, using the indoor and outdoor weather stations as well as temperature sensors attached to air-conditioning outlets.

For the En-Gage dataset, we used wearable sensors (i.e., **Empatica E4 wristbands**) as well as the same weather stations in the first longitudinal study. The two studies are located on the same campus, and the timelines of the two studies were partly overlapped. As a result, the collected data (i.e., weather information and occupant behaviours) in the longitudinal study can benefit the cross-sectional study or vice versa.

In the study, we tracked participants using Empatica E4 wristbands to measure physiological data, as well as daily surveys to query their thermal comfort, learning engagement and emotions while at school. Overall, we have collected 488 survey responses and 1415.56 hours of wearable data from all participants. During the data collection, one representative student was selected in each of the three Form classes. Their job was to distribute wristband sensors each morning, collect them after school and remind participants to complete the online surveys at the appropriate times. We anonymised the student's data by assigning each student an identity number (ID). Occupancy schedules were obtained from the individual classroom schedules provided by the school. These schedules can be used to represent the actual occupancy patterns of the building, although slight deviations from the planned schedule are to be expected in a school setting due to sickness and other circumstances.

Removal of Protected Health Information (PHI). We have de-identified the dates in the released data, and only kept the index of the week, weekday, day of the data collection, and time of the day. Specifically, we renamed the subfolders under the *Raw\_wearable\_data* folder using the index of week/weekday to indicate different days. We have also removed the date information in the survey responses, class tables and longitudinal data.

---

## Data Description

The data contains wearable sensor data (**electrodermal activity, heart rate, blood volume pulse, skin temperature, 3-axis acceleration data**), environmental data and self-report data (**student engagement, emotion, thermal comfort, seating behaviours**) collected from a field study in a K-12 high school.

The In-Gauge dataset consists of comma-separated variable (CSV) files - one for each classroom. Each classroom's spreadsheet contains time-related information and outdoor weather conditions (these are obviously identical for all classrooms). Furthermore, each classroom has information on its own indoor climate, whether or not it is occupied according to the class schedule, and information on whether its room air-conditioner is in heating or cooling mode.

For the En-Gage dataset, we have provided two versions: the original raw data by week/weekday and organised data based on the different class groups of the participants. The En-Gage dataset includes physiological signals measured using the Empatica E4 as well as self-reported data from the student and teacher participants.

Overall, the Longitudinal folder refers to the In-Gauge dataset and the other folders refer to the En-Gage datasets. Some useful notes for both datasets are provided below:

- **Longitudinal folder** contains all data pertaining to the longitudinal field study. It consists of 16 CSV files (one for each classroom). The CSV file names correspond to the classroom names. Each CSV file has a single header line and each of the following rows contains the following timestamped data at a resolution of 5 minutes per row.
- **Participant\_class\_info folder** contains demographic information on the background questionnaires and the *class* schedule. Note that for several survey questions, we adopted the 5-point Likert scale: -2 = 'strongly disagree', -1 = 'somewhat disagree', 0 = 'neither agree nor disagree', 1 = 'somewhat agree' and 2 = 'strongly agree'.
- **Survey folder** contains 2 files: *Student\_survey.csv* and *Teacher\_survey.csv*. *Student\_survey.csv* contains the 488 survey responses from student participants and *Teacher\_survey.csv* contains 22 survey responses by teachers.
- **Raw\_wearable\_data folder** includes 20 sub-folders named with the index of week and weekday (e.g., 'Week1\_4' indicates Thursday on Week 1), containing the raw wearable data for each day during the 4-week data collection. In each sub-folder, there are multiple sessions from different participants. Some participants provided more than 1 session on the same day. The name of each session consists of two parts connected by an underscore: a random string and the participant ID. For example, the session named 'kgarvl\_17' indicates the data is provided by participant 17. There are 6 CSV files in each session, and each of these (except *IBI.csv*) records the time and the sample rate expressed in Hz. It is worth noting that all the timestamps were removed except the time of the day.
- **Class\_wearable\_data folder** contains 221 sub-folders representing 221 different classes during which the wearable data were

recorded. Each sub-folder is named by the unique 'Class\_id' as shown in the Class\_table.csv. Each sub-folder includes further sub-folders named by the unique participant id or simply the label 'teacher'. These contain data from the wristband sensors for each participant in this class. There are 6 CSV files in each sub-folder: ACC.csv, EDA.csv, BVP.csv, HR.csv, IBI.csv, and TEMP.csv. The format of these files is identical to the ones in the *Raw\_wearable\_data* folder.

---

## Usage Notes

Our datasets include the outdoor/indoor/wearable sensing data and the self-report occupants' thermal comfort, learning engagement, and emotions while at school. This dataset is the first publicly available dataset for studying the daily behaviours and engagement of high school students using heterogeneous sensing. For the longitudinal outdoor and indoor sensing data, the most straightforward potential usage is to derive predictive models of how occupants operate room air-conditioning units. Our dataset could potentially be useful to examine the relationships between indoor/outdoor climates and physiological signals of occupants, which provide opportunities for the future design of intelligent feedback systems to benefit both students and staff on campus.

Specifically, various data mining (e.g., segmentation, clustering) and modelling techniques could be explored to build prediction models for measuring occupants' mental states using sensor-based physiological and behavioural recordings in buildings. This could be further used for various applications in future studies: (1) Monitoring signs of disengagement and negative emotions of students. Measuring the study engagement and emotions of students is beneficial to both teachers and students. Teachers will be able to improve their teaching strategies to create the right learning environment, improve the learning experience for students and re-engage students with low engagement. Students will be able to self-track their learning engagement and emotions, which could promote their self-regulation and reflective learning. (2) Studying peer effects in educational settings. It could be helpful to explore group-wise seating behaviours and their relationship to perceived engagement and physiological synchrony. (3) Providing comfortable indoor environments for occupants. It is possible to mitigate the negative effects of hot weather on student learning by using air conditioning, and teachers could ventilate classrooms timely to prevent excess carbon dioxide from affecting students' concentration.

---

## Release Notes

Initial release version 1.0.0 of the dataset.

The detailed data descriptor of In-Gauge and En-Gage can be found in Scientific Data [\[7\]](#). The dataset has been used in publications to predict student engagement in classes [\[3\]](#), analyse the classroom seating experience [\[6\]](#) and investigate the reliability of self-report surveys [\[4\]](#). Compared to previous versions [\[5\]](#), the current release has added information related to the week, weekday, and day of the data collection.

---

## Ethics

The data collection was approved by the Science, Engineering and Health College Human Ethics Advisory Network (SEH CHEAN) of RMIT University. The project was also approved by the principal of the school in which the study was conducted. Written informed consent was obtained from participants and guardians of minors prior to data collection.

---

## Acknowledgements

This research is supported by the Australian Government through the Australian Research Council's Linkage Projects funding scheme (project LP150100246 in partnership with Aurecon). This paper is also a contribution to the IEA EBC Annex 79.

---

## Conflicts of Interest

We declare no conflicts of interest.

---

## References

1. Haldi, F., & Robinson, D. (2011). The Impact of Occupants' Behaviour on Building Energy Demand. *Journal of Building Performance Simulation*, 4(4), 323-338.
2. Di Lascio, E., Gashi, S., & Santini, S. (2018). Unobtrusive assessment of students' emotional engagement during lectures using electrodermal activity sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(3), 1-21.
3. Gao, N., Shao, W., Rahaman, M. S., & Salim, F. D. (2020). n-Gage: Predicting in-class Emotional, Behavioural and Cognitive Engagement in the Wild. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(3), 1-26.
4. Gao, N., Saiedur Rahaman, M., Shao, W., & Salim, F. D. (2021, September). Investigating the Reliability of Self-report Data in the Wild: The Quest for Ground Truth. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and*

- Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers (pp. 237-242).
5. Gao, N., Marschall, M., Burry, J., Watkins, S., & Salim, F. (2021). In-Gauge and En-Gage Datasets. Figshare.
  6. Gao, N., Rahaman, M. S., Shao, W., Ji, K., & Salim, F. D. (2022). Individual and Group-wise Classroom Seating Experience: Effects on Student Engagement in Different Courses. arXiv preprint arXiv:2112.12342.
  7. Gao, N., Marschall, M., Burry, J. et al. Understanding Occupants' Behaviour, Engagement, Emotion, and Comfort Indoors with Heterogeneous Sensors and Wearables. Sci Data 9, 261 (2022). <https://doi.org/10.1038/s41597-022-01347-w>. Accessed on 9/6/2022.

## Contents

## Share



## Access

### Access Policy:

Anyone can access the files, as long as they conform to the terms of the specified license.

### License (for files):

[Open Data Commons Attribution License v1.0](#)

## Discovery

### DOI (version 1.0.0):

<https://doi.org/10.13026/srm3-7z33>

### DOI (latest version):

<https://doi.org/10.13026/zcpz-q137>

### Topics:

electrodermal activity heart rate physiological signals environmental sensing thermal comfort modelling occupant behaviour sensing emotion sensing human behavioural modelling smart building

### Project Website:

<https://www.nature.com/articles/s41597-022-01347-w>

## Corresponding Author

*You must be logged in to view the contact information.*


# Files

Total uncompressed size: 12.8 GB.

## Access the files

- [Download the ZIP file](#) (5.1 GB)
- Download the files using your terminal: `wget -r -N -c -np https://physionet.org/files/in-gauge-and-en-gage/1.0.0/`

Folder Navigation: <base>

Name		Size	Modified
 <a href="#">class_wearable_data</a>			
 <a href="#">longitudinal</a>			
 <a href="#">participant_class_info</a>			
 <a href="#">survey</a>			
 <a href="#">LICENSE.txt</a>		19.9 KB	2022-12-20
 <a href="#">README.txt</a>		11.2 KB	2022-08-25
 <a href="#">SHA256SUMS.txt</a>		669.4 KB	2023-02-13
 <a href="#">raw_wearable_data.zip</a>		1.5 GB	2022-12-09

PhysioNet is a repository of freely-available medical research data, managed by the MIT Laboratory for Computational Physiology.

Supported by the National Institute of Biomedical Imaging and Bioengineering (NIBIB) under NIH grant number R01EB030362.

For more accessibility options, see the [MIT Accessibility Page](#).

[Back to top](#)