scientific **data**

Check for updates

OPEN

DATA DESCRIPTOR

Understanding occupants' behaviour, engagement, emotion, and comfort indoors with heterogeneous sensors and wearables

Nan Gao [ID][1], Max Marschall[2], Jane Burry[3], Simon Watkins[4] & Flora D. Salim[5] ✉

We conducted a field study at a K-12 private school in the suburbs of Melbourne, Australia. The data capture contained two elements: First, a 5-month longitudinal field study *In-Gauge* 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. The dataset was used to derive predictive models of how occupants operate room air-conditioning units. Second, we tracked 23 students and 6 teachers in a 4-week cross-sectional study *En-Gage*, using wearable sensors to log physiological data, 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 & Summary

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]. One of the main factors driving this so-called 'performance gap' is the current misrepresentation of occupant behaviour in the simulations[2]. The software is accurate at modelling deterministic systems like automated air-conditioning units that are governed by set point temperatures, but incapable of accurately modelling the probabilistic nature of human behaviour, for example, the manual operation of air-conditioners. Occupant behaviour tends to be modelled on simplistic, rule-of-thumb assumptions that are not backed by data[3], usually by using the same set point approaches that are applied to

[1]School of Computing Technologies, RMIT University, Melbourne, 3000, Australia. [2]School of Architecture and Urban Design, RMIT University, Melbourne, 3000, Australia. [3]School of Design, Swinburne University of Technology, Melbourne, 3122, Australia. [4]School of Engineering, RMIT University, Melbourne, 3000, Australia. [5]School of Computer Science and Engineering, University of New South Wales (UNSW), Sydney, 1466, Australia. ✉e-mail: flora. salim@unsw.edu.au

Name	Year	Par.	Type	Modalities	Annotations	Duration	Scenario
Driving-stress[40]	2005	24	Field	ECG, EDA, EMG, RESP	Stress level	>50 minutes	Real-world driving tasks
DEAP[41]	2011	32	Lab	Videos, EEG, EDA, BVP, RESP, ST, EMG and EOG	Arousal, valence, like/dislike, dominance, familiarity	40 minutes	Watch music videos
Driving-work[42]	2013	10	Field	EDA, HR, TEMP	Mental workload	30 minutes	Drive a predefined route
StudentLife[43]	2014	48	Field	Smartphone	Stress, mood, happiness	10 weeks	Real life, student exams
DECAF[44]	2015	30	Lab	ECG, EMG, EOG, MEG, near-infrared face, video	Valence, arousal, and dominance	>1 hour	Watch music video and movie clips
Non-EEG[45]	2016	20	Lab	ACC, EDA, HR, TEMP, SpO2	N/A	<1 hours	Four types of stress (physical,emotional, cognitive, none)
Ascertain[46]	2016	58	Lab	ECG, EDA, EEG, facial features	Arousal, valence, engagement, liking, familiarity, personality	90 minutes	Watch movie clips
Stress-math[47]	2017	21	Lab	ACC, EDA, HR, TEMP	Anxiety	26 hours (total)	Solve math questions under different pressure
WESAD[48]	2018	15	Lab	ACC, BVP, ECG, EDA, EMG, RESP, TEMP	Affect, anxiety, stress	2 hours	Neutral, amusement and stress conditions
Snake[49]	2020	23	Lab	ACC, BVP, EDA, TEMP	Cognitive load, personality	>6 minutes	Smartphone games with three difficulty levels
CogLoad[49]	2020	23	Lab	ACC, BVP, EDA, TEMP	Cognitive load, personality	N/A	6 cognition load tasks
K-EmoCon[50]	2020	32	Lab	Videos, audio, ACC, EDA, EEG, ECG, BVP, TEMP	Arousal, valence, stress, affect	173 minutes (total)	Social interaction scenario involving two people
En-Gage	2022	29	Field	ACC, EDA, BVP, TEMP, In. TEMP, HUMID., CO2, NOISE	Cognitive, behavioural, emotion engagement, thermal comfort, arousal, valence	4 weeks (1416 hours in total)	Real-world courses in a high school

Table 1. Publicly available datasets in the affective computing area.

automated systems (e.g. occupant switches on the air-conditioner when the indoor temperature exceeds 24 °C). Actual human behaviour is less responsive and more varied; thus, researchers have conducted field studies in operational buildings, by measuring various environmental and other variables alongside an observed behaviour (for example, the operation of air-conditioners, windows, lights, fans, etc.). They use this data to derive statistical models of the observed behaviour based on one or several of the observed independent variables[4–6]. The first assay of our study contributes data towards this endeavour, specifically enabling the creation of predictive models of occupants' use of room air-conditioners in schools.

The second assay is a four-week cross-sectional study tracking 23 students and six teachers, using wearable sensors to log physiological data, as well as daily surveys to query the occupants' thermal comfort, learning engagement, seating locations and emotions while at school. Buildings contribute about a third of world energy consumption, mainly due to the use of heating, ventilation and air-conditioning (HVAC) systems for indoor climate regulations. Considering that we have spent so much energy and effort in providing adequate environments for building occupants, it is worth investigating what exactly constitutes their comfort and well-being. The above-mentioned building performance simulations tend to define comfort either by using deemed-to-satisfy temperature thresholds or by using comfort models, most commonly the predicted mean vote (PMV) model. However, the PMV model has not been updated since it was derived from laboratory experiments in the 1960s. It has been criticised for its poor predictive performance in real-world contexts[7] and does not appear to apply to all age groups[8]. Furthermore, thermal acceptance is clearly only one of several metrics for assessing indoor well-being.

On the other hand, 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[9,10]. 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 and engagement levels. For example, EDA is generally considered to be a good indicator of psychological arousal and has been increasingly studied for the detection of engagement[9,10], emotion[11], and depression[12], etc. 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). Table 1 shows how *En-Gage* dataset is distinguished from existing emotion datasets.

En-Gage is the first publicly available dataset studying the daily behaviours and engagement of high school students using heterogeneous sensing. Together with *In-Gauge* dataset, it offers a unique opportunity to analyse the relationships between indoor climates and the mental states of school students - not only related to their thermal

Group	Room	Participant
Form	R1	P13, P14, P15, P16, P17, P18, P19, P20, P21, P22
	R2	P8, P9, P10, P11, P12, P23
	R3	P1, P2, P3, P4, P5, P6, P7
Maths	R1	P2, P4, P5, P10, P11, P14, P18
	R2	P3, P6, P7, P8, P9, P15, P16, P17, P20
	R3	P1, P12, P13, P19, P21, P22, P23
Language	R1	P1, P2, P4, P7, P10, P13, P15, P17, P19, P20, P21, P22, P23
	R2	P9, P14
	R3	P5, P6, P11, P12, P16
	R4	P3, P8 P18

Table 2. Distribution of student participants in different class groups.

comfort but also their emotions, engagement and productivity while at school. Especially, it's unusual to combine individual sensor data with building environmental data together, to study how indoor and outdoor environments influence the complex occupant behaviours and physiological responses. The combination of two datasets will benefit building scientists, behaviour psychologists and affective computing researchers in future research.

Methods

Ethics approval. The data collection was approved by the Science, Engineering and Health College Human Ethics Advisory Network (SEH CHEAN) of RMIT University. SEH CHEAN also reviewed and approved the consent forms for participants and guardians of minors, which included information on the purpose of and procedures for the research, the types of data to be collected, the compensation for the involvement and the protocols for privacy protection and data storage. The project was also approved by the principal of the school in which the study was conducted.

Participants and recruitment. For the cross-sectional study, we recruited participants from a K-12 private school in the suburbs of Melbourne (pop. 700). The recruitment occurred in August and September 2019, and calls for participation were disseminated through information leaflets, recruitment letters and a presentation in the school hall, with the assistance of the director teacher of the Year 10 students (Year 10 is the eleventh year of compulsory education in Australia). The admission was restricted to Year 10 students and their teachers whose native language was English or who were bilingual. A total of 23 (15–17 years old, 13 female and 10 male) out of 75 Year 10 students and six (33–62 years old, four female and two male) out of 12 teachers met the inclusion criteria, volunteered for the study and signed the consent forms. Since all the student participants were underage, their guardians also provided signed consent forms. Raw data for $n = 23$ student participants were properly recorded and nearly complete (but with different wristband wearing days), constituting the majority of the *En-Gage* dataset.

The volunteers were then asked to complete an online background survey, which was accessible through a web page link that was shared with them. In the survey, we collected information on the participants' age, gender, general thermal comfort and classes. The Year 10 students at the school were taught in separate class groups. They were separated into three *Form* groups for English, Science, Global Politics, Physical Education and Health/Sport courses, three *Maths* groups and four *Language* groups (see Table 2). Asking for each student's class group in the background survey allowed us to determine which classroom they were in at any given time. Among the participating teachers, there were three math teachers, one English teacher, one Japanese teacher, and one science teacher.

As a token of appreciation for their participation, we awarded each participating student with a certificate of participation and four movie vouchers - one for each week of successful participation. Participation in this research project was voluntary, and we communicated to participants that they were free to withdraw from the project at any stage.

Experiment setup. We conducted our study at a mixed-gender K-12 private school. The longitudinal study was conducted for a 5.5-month period from 7 October 2019 to 23 March 2020, using the indoor and outdoor weather stations as well as temperature sensors attached to air-conditioning outlets. The cross-sectional study included four weeks of data capture: the first two weeks of data collection started from 2 September 2019, and the second two weeks of data collection started from 28 October 2019, using the wearable sensors 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. For instance, the outdoor temperature and humidity can help researchers understand student clothing insulation and thermal comfort on campus. Additionally, the physiological signals can be combined with the environmental signals to accurately predict the heating/cooling behaviours of occupants in buildings.

In the study, we tracked participants using *Empatica E4* wristbands (see Fig. 1(a)) 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
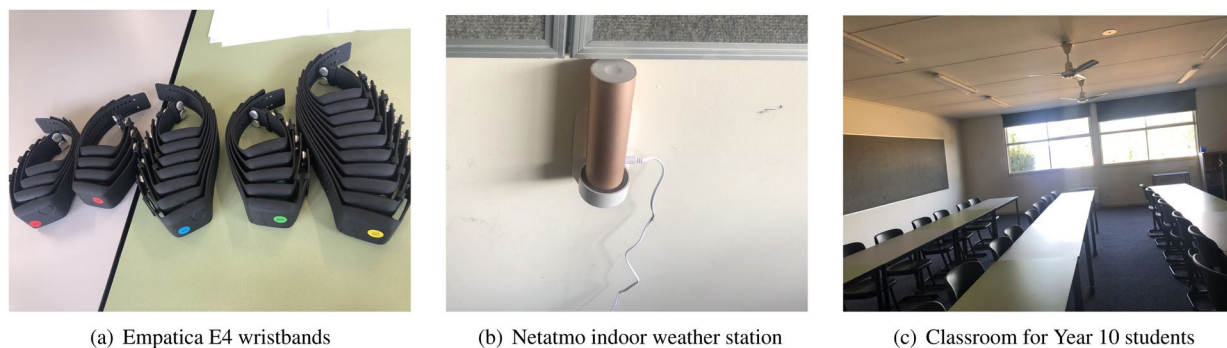
| (a) Empatica E4 wristbands | (b) Netatmo indoor weather station | (c) Classroom for Year 10 students |

Fig. 1 Devices and environments for collecting wearable and indoor data.

Devices	Collected data	Sampling rate	Time frame
Empatica E4 wristband	3-axis acceleration	32 Hz	4 weeks
	Skin temperature	4 Hz	
	Electrodermal activity	4 Hz	
	Blood volume pulse	64 Hz	
Netatmo indoor weather station	Humidity, temperature, noise level, CO2	5 minutes	5.5 months
DigiTech XC0422 outdoor weather station	Temperature, humidity, barometric pressure, wind speed, wind direction, solar radiation, UV, rainfall	5 minutes	5.5 months
PHILIO Z-wave (attached to air-conditioning vents)	Humidity, temperature	5 minutes	5.5 months

Table 3. Data collected with sensors with respective sampling rate and time.

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. The following is a description of the research instruments used in the study (see Table 3).

Daily surveys. On each school day, student participants were asked to complete online surveys (either through tablets placed in each classroom or using their own digital devices) at 11:00, 13:25 and 15:35 (directly after the second, fourth and fifth class. The length of the second and fourth class was either 40 min or 80 min, depending on the day of the week, and the fifth class always lasted 80 min. The curriculum in this school had a bi-weekly rhythm, i.e., the first and second weeks had different class schedules, but the first and third weeks were identical, as were the second and fourth weeks. The representative student was tasked with reminding the student participants to complete the online surveys on time, as described in Table 4. The online questionnaire included 11 items related to the students' psychological states and behaviours (e.g., thermal comfort, student engagement and emotions). All the items (except the seating location and confidence level) were used directly or slightly adapted from the validated questionnaires widely used by researchers in this area. The screenshot of the question for seating location can be seen from Fig. 2(a). Figure 2(b) illustrates the implemented PAM[13] asking the user to choose one picture from a grid of 16 pictures in a library of 32 photos. Figure 3 displays the distribution of responses to the thermal sensation (from −3 to 3), thermal preference and clothing level. The distribution of multidimensional (behavioural, emotional and cognitive) engagement can be found in Fig. 4(a), and the overall engagement across participants is reflected in Fig. 4(b). Figure 3(c) depicts the distribution of emotions in the valence and arousal dimensions. The numbers indicate the percentage frequencies, and the darker the colour, the higher the frequency of the specific emotion (e.g., arousal = 1 and valence = 2). Figure 5 shows the distribution of seating locations across different participants.

Figure 6(a) shows the distribution of survey responses throughout the day. As students were requested to submit their self-reports directly after the second, fourth, and fifth classes (i.e., 11:00, 13:25 and 15:35, respectively), most responses were submitted at 11:00–12:00, 13:00–14:00 and 15:00–16:00. The survey responses that were recorded before the start of the targeted class or after the start of the next class should be removed if we aim to explore student engagement in the targeted class. However, if we wish to study thermal comfort, clothing insulation, emotion or confidence level, all the survey responses can be kept. Next, we explored the distribution of the survey responses through the week (see Fig. 6(b)). We found that most students submitted their responses on Monday or Thursday. The number of students who submitted self-reported data on Wednesday was the lowest. The potential reason for this may be that students forgot to submit their responses, as they took breadth studies on Wednesday (normal studies on the other weekdays).

Annotation categories	Description	Measurement scale
Thermal sensation	Commonly used ASHRAE thermal sensation[51]	−3: cold, −2: cool, −1: slightly cool, 0: neutral, 1: slightly warm, 2: warm, 3: hot
Thermal preference	Commonly used ASHRAE thermal preference[51]	Choose one (cooler, no change, warmer)
Clothing level	Commonly used ASHRAE clothing insulation[51]	Choose multiple
Seating location	Seating location in the classroom	Click one point
Behavioural/Emotional/Cognitive engagement	Adapted In-class Student Engagement Questionnaires (ISEQ)[52]	−2: disagree, −1: somewhat disagree, 0: neither agree nor disagree, 1: somewhat agree, 2: strongly agree
Arousal/Valence	Commonly used affective dimensions from the Photographic Affect Meter (PAM)[13]	Choose one photo
Confidence level	Confidence level of the response	1: not confident, 2: slightly confident, 3: moderately confident, 4: very confident, 5: extremely confident

Table 4. Collected annotations from the questionnaires.
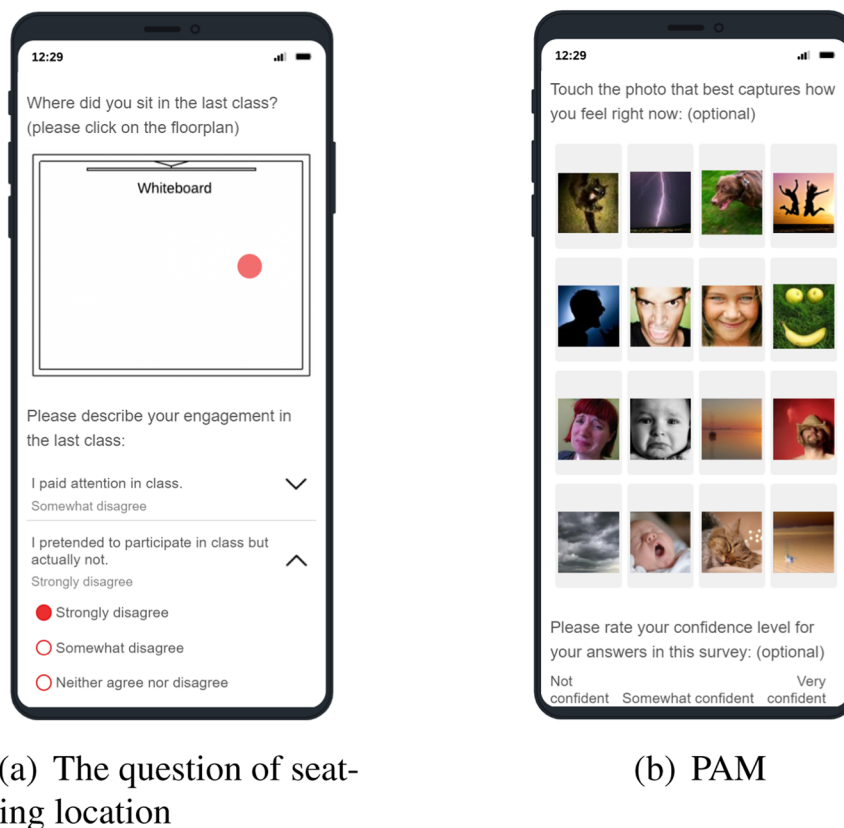
(a) The question of seating location

(b) PAM

Fig. 2 Screenshots of the self-report survey.

Empatica E4 wristband. These wristband sensors (see Fig. 1(a)) were first proposed for use in studies by Garbarino *et al.*[14]. These watch-like devices have multiple sensors: an EDA sensor, a photoplethysmography (PPG) sensor, a three-axis accelerometer (ACC) and an optical thermometer. EDA refers to constantly fluctuating changes in the electrical properties of the skin at 4 Hz; when the level of sweat increases, the conductivity of the skin increases. PPG sensors measure the blood volume pulse (BVP) at 64 Hz, from which the interbeat interval (IBI) and heart rate variability (HRV) can be derived. The ACC records in the range of $[-2\,g, 2\,g]$ at 32 Hz and captures motion-based activity, which has been widely used in smartphones, wearables and other IoT devices[15]. The optical thermometer reads peripheral skin temperature (ST) at 4 Hz. In recording mode, E4 wristbands can store 60 hours of data in memory, with a battery life of over 32 hours. They are lightweight, comfortable and waterproof, and were thus especially suitable for the continuous and unobtrusive monitoring of the participants in our study. Before the data collection, all wristbands were synchronised with the E4 Manager App, using a single laptop to ensure that the internal clocks were accurate. Each student was assigned a wristband sensor marked with their unique study ID. The students were asked to wear the wristband on their non-dominant hand, and to avoid pressing the button or performing any unnecessary movements during class. The teacher participants were only required to wear the wristbands while teaching the year 10 classes. Figure 7 displays the
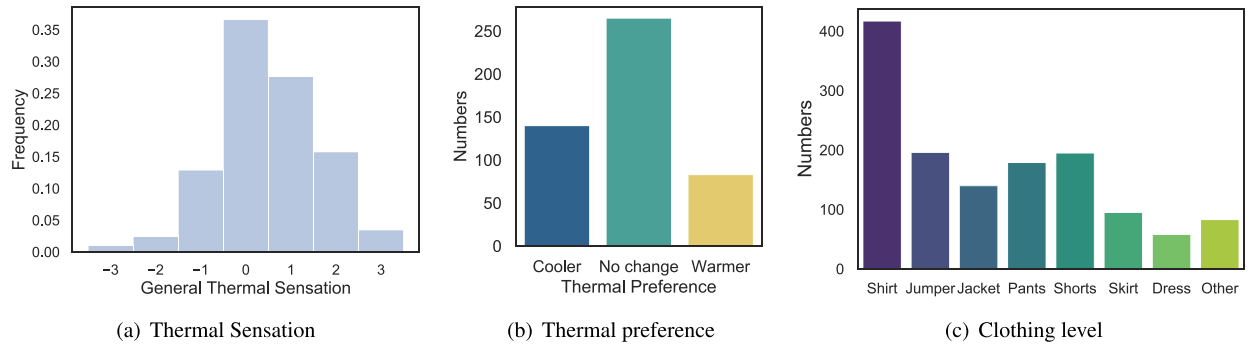
(a) Thermal Sensation   (b) Thermal preference   (c) Clothing level

Fig. 3 Distribution of responses related to thermal comfort.

(a) Multi-dimensional Engagement   (b) Overall Engagement across Participants   (c) Valence and Arousal
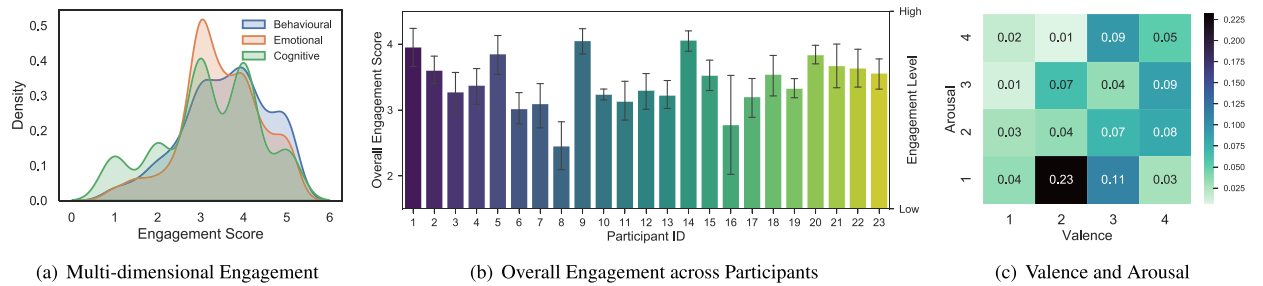
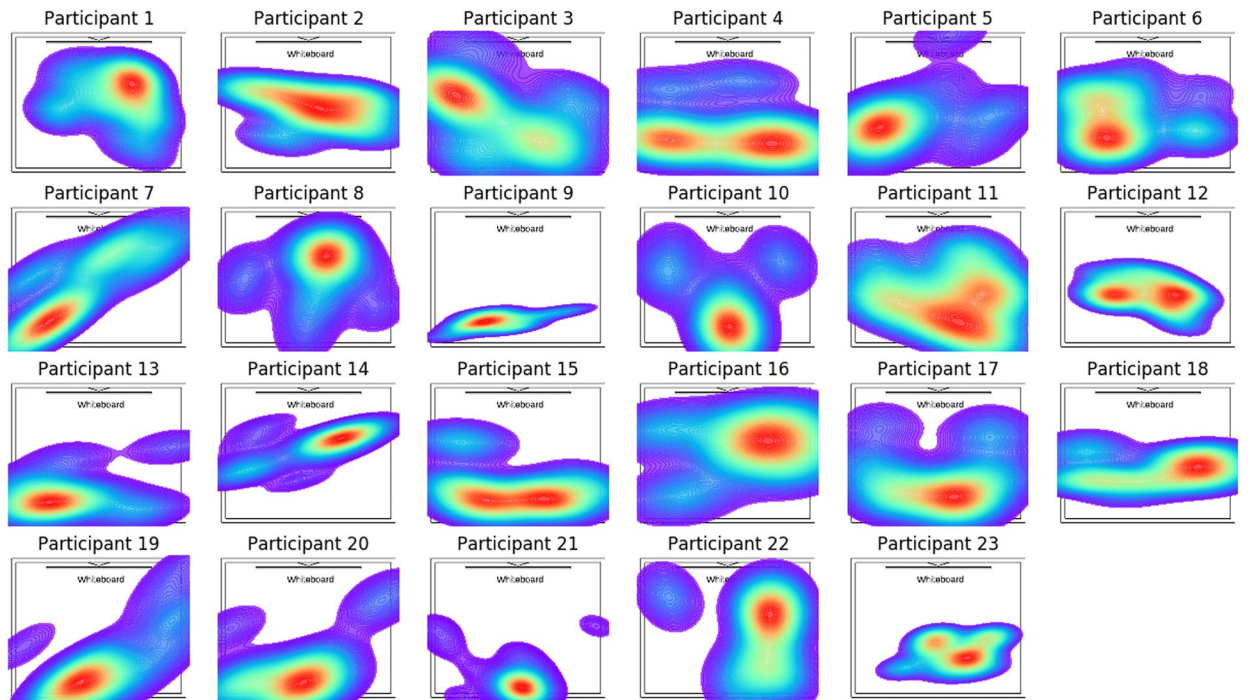**Fig. 4** Distribution of responses related to the engagement and emotion.

Fig. 5 Distribution of seating locations across different participants.

distribution of wearable signals per school day for all participants. The blue line indicates the average values of signals calculated from 369 traces during school time (9:00 to 15:30).

DigiTech XC0422. We set up two outdoor weather stations on-site: one in the prevailing NNW windward direction located at some distance from the buildings, and one on the SSE leeward side. These logged the data types shown in Table 5 at 5-minute intervals via the school's guest WiFi to WUnderground.com where it can be
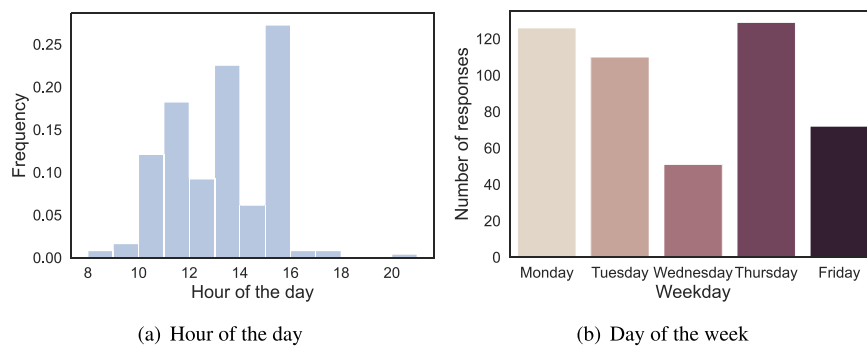
(a) Hour of the day   (b) Day of the week

Fig. 6 Distribution of the survey responses over hours of the day or day of the week.

accessed remotely. Note that these weather stations log solar irradiance values in W/m2 but only have a luminosity sensor. The method of conversion from lux to W/m2 is unclear from the product's datasheet, but we assumed that it was in line with a commonly used, simplified conversion rate (e.g. Michael, 2019)[16].

Netatmo healthy home coach. We collected indoor environmental data using Netatmo Healthy Home Coaches installed in 17 classrooms as shown in Fig. 1(b,c). The logging specifications of these devices are shown in Table 6. These devices measure indoor temperature, relative humidity, CO_2 levels and noise levels at a five-minute logging frequency. The data was uploaded in real-time via the school's guest WiFi to the Netatmo cloud platform from which we accessed the data remotely through our Netatmo account login. The analysed classrooms differed from one another in several aspects including the room geometry and orientation, as well as the number and location of windows. The placement of environmental sensor devices was therefore determined on a case-by-case basis, with the goal of finding an optimal trade-off between several, partly conflicting considerations, most of which were suggested by Wagner *et al.*[17]. For example, we tried placing the sensors close to the occupants but at the same time avoiding the sensors from being obstructed, biased or obtrusive due to their proximity to the occupants, furniture, heating elements, vents or appliances. The ASHRAE Standard 55 recommends temperature sensor heights of 0.1, 0.6 and 1.1 m for ankles, waists and heads of seated occupants, respectively. Given these guidelines, since only one device was installed per room in this study and the head height of children is lower than that of adults, we attempted to place the sensors at approximately 0.9 m. Three of the classrooms with a Netatmo station were rooms frequented by Year 10 students, therefore this data could be used in combination with the data captured by the E4 wristband sensors in the cross-sectional study. It should be noted that we started using these three devices and the outdoor weather stations before beginning the full longitudinal field study in all classrooms.

Philio temperature/humidity sensor. The classrooms had split-system remote-controlled air-conditioning units for heating and cooling. We inferred their usage by measuring temperature fluctuations with Philio Temperature/Humidity Sensors placed at the outlets of the vents of the remote-controlled room air conditioning units. The sensors logged data via Z-Wave to Vera Edge hubs, several of which were placed throughout the school due to their limited range. Data was logged at 5-minute intervals using custom LUA scripts via the VeraAlerts app within the Vera SmartHome app, which enabled sending the data to the Pushbullet online platform from where they could be accessed remotely.

Data post-processing. For the longitudinal study, we extracted the environmental data from their respective online platforms and rounded each data point's timestamp to the nearest 5-minute step to enable the aggregation of data from different sources, and interpolated over missing data points. The data from the two weather stations were averaged for each time step. In cases where one of the stations had missing data, we used the other weather station's data point. The outdoor wind direction was originally given in a 16-step scale of cardinal directions which we converted to numerical angle values in degrees.

Within the context of this field study, there was no way to directly monitor when the air conditioning units were in use. Instead, we measured their use indirectly with the Philio temperature sensors mounted to the air conditioning outlets. Creating an algorithm that reliably distinguishes all four event types (cooling switched on, cooling switched off, heating switched on and heating switched off) is a task that would have exceeded the scope of our research. Instead, we used threshold values of the temperature slope to predict events. If the current state was off, then a sudden rise would be classified as switching on the heating; if the current state was cooling, the same rise in temperature would be classified as switching off the cooling. We found that classifying a temperature difference of 0.5 degrees from one 5-minute time step to the next to be useful in auto-detecting the majority of switching events. However, this crude method was limited in its predictive capability, not least because in an 'on' state, an air-conditioner will automatically keep switching itself off and on in order to remain within a certain temperature band around the selected set point; an example of this can be seen in the last air-conditioning period shown in Fig. 8. Therefore, we relied on a visual assessment of the data and manually overwrote time frames with states that appeared to have been incorrectly categorised by the algorithm, based on the curves of the indoor temperature and AC temperature graphs. This is a potential source of error, but we assumed that the
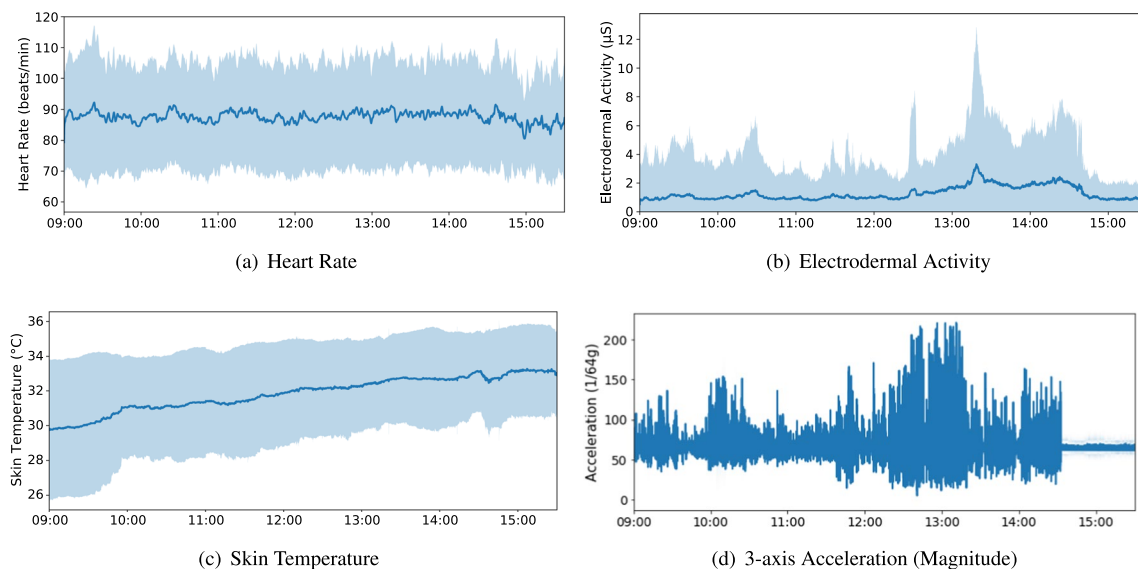
(a) Heart Rate

(b) Electrodermal Activity

(c) Skin Temperature

(d) 3-axis Acceleration (Magnitude)

Fig. 7 Wearable signals per school day for all participants (369 traces in total).

Type	Units	Range	Accuracy	Resolution
Dry bulb temperature	°C	−40°C–60°C	±1%	0.1°C
Dew point temperature	°C	−40°C–60°C	±1%	0.1°C
Relative humidity	%	1%–99%	±5%	1%
Wind speed	m/s	0 m/s–50 m/s	±1 m/s (<5 m/s)	0.1 m/s
			±10% (≥5 m/s)	
Gust speed	m/s	0 m/s–50 m/s	±1 m/s (<5 m/s)	0.1 m/s
			±10% (≥5 m/s)	
Wind direction	°	0°–360°	±22.5°	22.5°
Rainfall	mm	0 mm–9,999 mm	±10%	0.3 mm (<1000 mm)
				1 mm (≥1000 mm)
Light	lux	0 lux–400,000 lux	±15%	0.1 lux
Solar radiation	W/m^2	N/A	N/A	N/A

Table 5. DigiTech XC0422 logging specifications.

Type	Units	Range	Accuracy	Resolution
Dry bulb temperature	°C	0°C–50°C	±0.3°C	0.1°C
Relative humidity	%	0%–100%	±3%	1%
CO$_2$	ppm	0 ppm–5,000 ppm	±50 ppm (<1,000 ppm)	1 ppm
			±5% (≥1,000 ppm)	
Noise	dB	35 dB–120 dB	N/A	1 dB

Table 6. Netatmo Healthy Home Coach logging specifications.

assessment was sufficiently accurate for this study - an assumption that proved correct when testing it on site. We aggregated all the data into spreadsheets for each classroom individually, and added several data, including columns that identified holiday periods, occupancy and time frames of insufficient data coverage.

In the cross-sectional study, for the wearable data, we converted the timestamps of the wristband sensor readings from raw time intervals and Unix time to the local date-time format. Then we categorized the wearable data based on a different date. The wearable data were extracted according to the scheduled length for the second, fourth and fifth class which ends at 11:00, 13:25, 15:35. For the online survey, we received a total of 488 valid online surveys from students with a response rate of 35.3%. We also received 22 online surveys from teachers. Then, we aligned the survey data to one of the three classes. We set the survey responses before 11:25 pm to belong to the second class, between 12:15 pm–14:15 pm to belong to the fourth class, after 14:15 pm to belong to the fifth class.
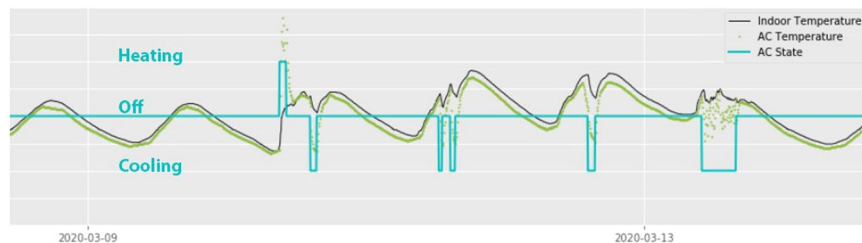
Fig. 8 Data sample showing the indoor ambient temperature, the temperature reading at the air conditioning vent and the inferred air conditioning states.

Data Records

Summary.   The data are available on the *Figshare* data repository[18]. It includes two elements: *In-Gauge* dataset for the longitudinal study and *En-Gage* dataset for the cross-sectional study. For the *En-Gage* dataset, we have provided two versions: the original raw data by date and organised data based on the different class groups of the participants.

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. The *En-Gage* dataset includes physiological signals measured with the wristband sensors as well as self-reported engagement, thermal comfort, seating locations, and emotion data from the student and teacher participants.

Contents.   In the following, we describe the directories and files in our datasets.

Longitudinal.   This folder contains all data pertaining to the longitudinal field study. It consists of a TXT file describing the dataset and 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:

- Timestamp: Local datetime format e.g. '2019-10-08 18:25:00'.
- Year: An integer of either 2019 or 2020.
- Month: An integer between 1 and 12.
- DayOfYear: An integer between 1 and 365.
- Occupied: '0' means that the room was not occupied at this time according to the classroom schedule; '1' means it was.
- SchoolDay: '0' means that this day was not a school day; '1' means it was.
- Hour: An integer representing the hour od day from 0 to 23.
- LessonNumber: An integer signifying which class is currently taking place (note that each school day started with a 10-minute assembly referred to here as the '0'th class):'-1' = outside of school hours; '0' = 8:50–9:00; '1' = 9:00–9:40; '2' = 9:40–10:20; '3' = 10:20–11:00; '4' = 11:25–12:05; '5' = 12:05–12:45; '6' = 12:45–13:25; '7' = 14:15–14:55; '8' = 14:55–15:35; '9' = Recess times or special "Breadth Studies" session on Wednesdays.
- LessonPct: A fraction between 0.0 and 1.0 describing how much of the current class has passed.
- IndoorTemperature: A decimal number representing the current indoor temperature in °C.
- IndoorHumidity: An integer representing the current indoor relative humidity in %.
- IndoorCO2: An integer representing the current indoor CO_2 concentration in ppm.
- IndoorNoise: An integer representing the current indoor noise level in dB.
- OutdoorTemperature: A decimal number representing the current outdoor temperature in °C.
- OutdoorHumidity: An integer representing the current outdoor relative humidity in %.
- OutdoorDewpoint: A decimal number representing the current outdoor dewpoint temperature °C.
- OutdoorWindDirection: An integer representing the current outdoor wind direction in degrees, from 0 to 360 (0° = north wind, 90° = east wind, etc.).
- OutdoorWindSpeed: A decimal number representing the current outdoor wind speed in m/s.
- OutdoorGustSpeed: A decimal number representing the current outdoor gust speed in m/s.
- Precipitation: A decimal number representing the current outdoor precipitation in mm.
- UvLevel: An integer between 0 and 11 representing the current outdoor Global Solar UV Index.
- SolarRadiation: An integer representing the current outdoor solar radiation intensity in W/m².
- CoolingState: '0' means that the room air-conditioner was currently not cooling the room; '1' means it was.
- HeatingState: '0' means that the room air-conditioner was currently not heating the room; '1' means it was.
- UsabilityMask: For timeframes where too much data was missing, we set this UsabilityMask field to "False" for the entire day. During holidays, the UsabilityMask also reads "False".

Participant_class_info. This folder contains demographic information on the background questionnaires participants, 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'. The *Participant_class_info* folder contains the following files:

1. *Student.csv.* Each row in this file contains a participant ID (*Column A*), gender (*Column B*), age in years (*Column C*), form room, math room and language room (*Columns D–F*), and three background questions (*Columns G–K*) related to their general thermal comfort and engagement in class. Specifically, *Columns G to I* represent, respectively, the questions 'What is your general feeling in the classroom?' [−3 = cold, −2 = cool, −1 = slightly cool, 0 = neutral, 1 = slightly warm, 2 = warm, 3 = hot], 'When I am engaged in class, I usually don't feel too hot or too cold' and 'When I am engaged in class, I could get distracted when the room is too hot or too cold'. For the latter 2 questions, we adopted the 5-point Likert scale.

2. *Teacher.csv.* Each row in this file contains a participant ID (*Column A*), gender (*Column B*), age in years (*Column C*), teaching subject (*Columns D*), and three background questions similar to the *student.csv* file, except that we changed the last two questions slightly from 'When I am engaged in class, […]' to 'When I am engaged in teaching, […]'.

3. *Class_table.csv.* We generate this file from the class schedule obtained from the school. Each row in this file contains the information of one single class, including the unique class ID (*Column A*), classroom (*Column B*), date (*Column C*), start time of the current class (*Column D*), finish time of the current class (*Column E*), length of the class (*Column F*), week (*Column G*), weekday (*Column H*), the order of the class (*Column I*) and the course name (*Column J*). Specifically, *Column K* shows whether students take this class in a form group, where '0' indicates they are not in a form group, 'all' indicates all students take this class in one whole form group (i.e., Assembly, Chapel), the R1/R2/R3 means students take this class in form groups and their form room is R1, R2 or R3.

Survey. This folder contains 2 files: *Student_survey.csv* and *Teacher_survey.csv.*

Student_survey.csv contains the 488 survey responses including 15 columns where *Column A* is participant ID and *Column B* is the recorded time. There are columns containing thermal comfort-related information (*Columns C–G*), multi-dimensional student engagement (*Columns H–L*), mood (*Column M*), and confidence level of the survey (*Column N*). The engagement questions were rated using the Likert-scale. To calculate the engagement score, users should reverse the responses in item 2 and item 4, then calculate the average of the 5-point Likert scale for each dimension of engagement. The specific columns relate to the following questions:

- *Column C: Thermal_sensation*: "How do you feel right now in the classroom?" [−3 = cold, −2 = cool, −1 = slightly cool, 0 = neutral, 1 = slightly warm, 2 = warm, 3 = hot].
- *Column D: Thermal_preference*: "Would you like to be?" [Cooler, No change, Warmer].
- *Column E: Clothing*: "What are you wearing now? (multiple options allowed)" [Shirt, Jumper, Jacket, Pants, Shorts, Skirt, Dress, Other].
- *Columns F–G: Loc_x, Loc_y*: "Where did you sit in the last class? (please click on the floorplan)" [x, y pixels in the 400*321 room thumbnail where x = y = 0 at the upper left corner].
- *Columns H–L: Engage_1, 2, 3, 4, 5*: "Please describe your engagement in the last class": [I paid attention in class], [I pretended to participate in class but actually not], [I enjoyed learning new things in class], [I felt discouraged when we worked on something], [I asked myself questions to make sure I understood the class content].
- *Columns M–N: Arousal, Valence*: "Touch the photo that best captures how you feel right now (optional)" [We assigned the arousal and valence values from 1–4 to each picture. For instance, for the right bottom picture, valence = 4 and arousal = 1].
- *Column N: Confidence_level*: "Please rate your confidence level for your answers in this survey (optional)" [5-point Likert scales where 1 = Not confident, 2 = Slightly confident, 3 = Moderately confident, 4 = Very confident, 5 = Extremely confident].

Teacher_survey.csv contains the 22 survey responses by the teachers. The file includes 11 columns where *Column A* is the recorded time, *Column B* is the wristband ID, *Columns C–E* are the thermal comfort-related information, *Columns F–G* are the engagement related information, and *Column K* is the confidence level of the survey. For the wristband ID in *Column B*, A/B/C/D represent the classrooms R1/R2/R3/R4. The specific columns relate to the following questions:

- *Column B: Wristband_id*: "Please enter your wristband ID." [A, B, C, D].
- *Column C: Thermal_sensation*: "How do you feel right now in the classroom?" [−3 = cold, −2 = cool, −1 = slightly cool, 0 = neutral, 1 = slightly warm, 2 = warm, 3 = hot].
- *Column D: Thermal_preference*: "Would you like to be?" [Cooler, No change, Warmer].
- *Column E: Clothing*: "What are you wearing now? (multiple options allowed)" [Shirt, Jumper, Jacket, Pants, Shorts, Skirt, Dress, Other].
- *Columns F–G: Engage_1, 2, 3, 4, 5*: "Please describe your engagement in the last class": [I was excited about teaching], [I felt happy while teaching], [While teaching, I paid a lot of attention to my work], [I cared about the problems of my students], [I was aware of my students' feelings].
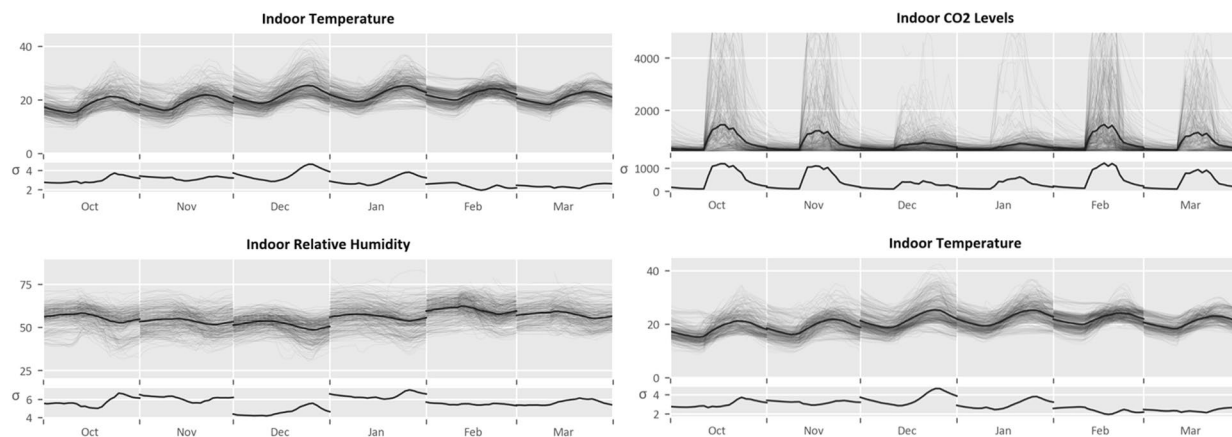
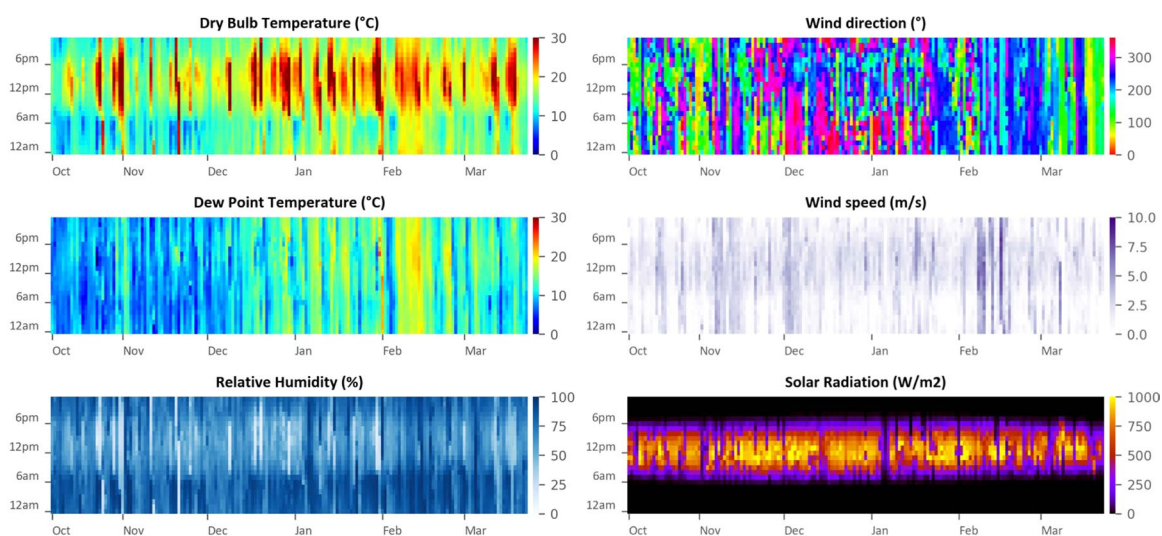**Fig. 9** Daily indoor environmental trends by month.

Fig. 10 Hourly outdoor climate (averaged between the two weather stations).

- *Column K: Confidence_level*: "Please rate your confidence level for your answers in this survey (optional)" [5-point Likert scales where 1 = Not confident, 3 = Somewhat confident, 5 = Very confident].

Raw_wearable_data. This folder includes 20 sub-folders named with the date of data collection (e.g., '20191122'), 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: the unique session ID and the participant ID. For example, the session named '1567380164_18' indicates the data is provided by participant 18. There are 6 *CSV* files in each session, and each of these files (except *IBI.csv*) has the following format: the first row is the initial time of the session expressed as a Unix timestamp in UTC. The second row is the sample rate expressed in Hz. Specifically:

1. *ACC.csv* contains data from a 3-axis accelerometer sampled at 32 Hz which is configured to measure accelerations in the range of $[-2g, 2g]$. Acceleration is the rate of change of the velocity with respect to time, where SI (International System of Units)[19] derived unit for acceleration is the metre per second squared $(m \cdot s^{-2})$ where 1 g is equal to 9.80665 $m \cdot s^{-2}$. The unit in this file is 1/64 g where the raw value of 64 indicates 1 g. The 3 columns refer to the x, y, and z-axis, respectively.
2. *BVP.csv* contains BVP signals sampled at 64 Hz which is the primary output from the PPG sensor. BVP signals can be used to compute the inter-beat-intervals (IBI) and heart rate (HR).
3. *EDA.csv* contains data from an electrodermal activity (EDA) sensor expressed as micro siemens (μS) sampled at 4 Hz. The variation of EDA values indicates the electrical changes of the skin surface and the EDA arises when the skin receives nerve signals from the brain and sweat level increases[20].
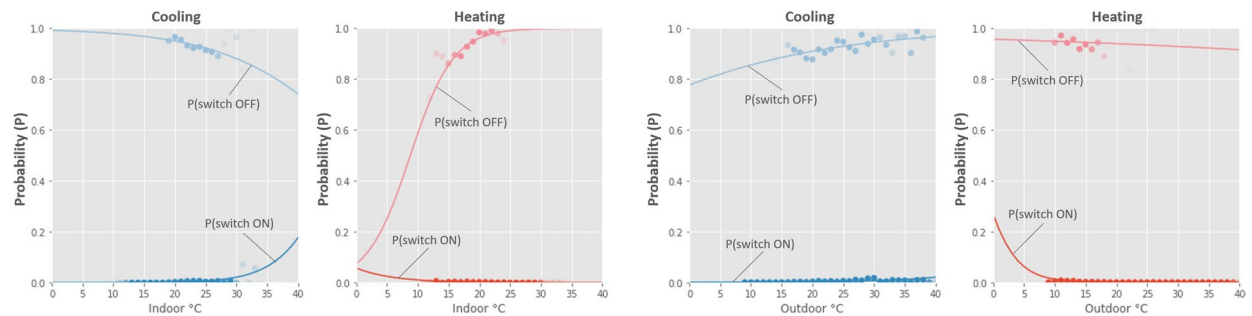
Fig. 11 An example for an application of the dataset: creating models of occupant behaviour to predict the switching of air-conditioners, for both heating and cooling. The above are examples using simple logistic regression; the left ones use indoor temperature as the independent variable, the right ones use outdoor temperature as the independent variable.
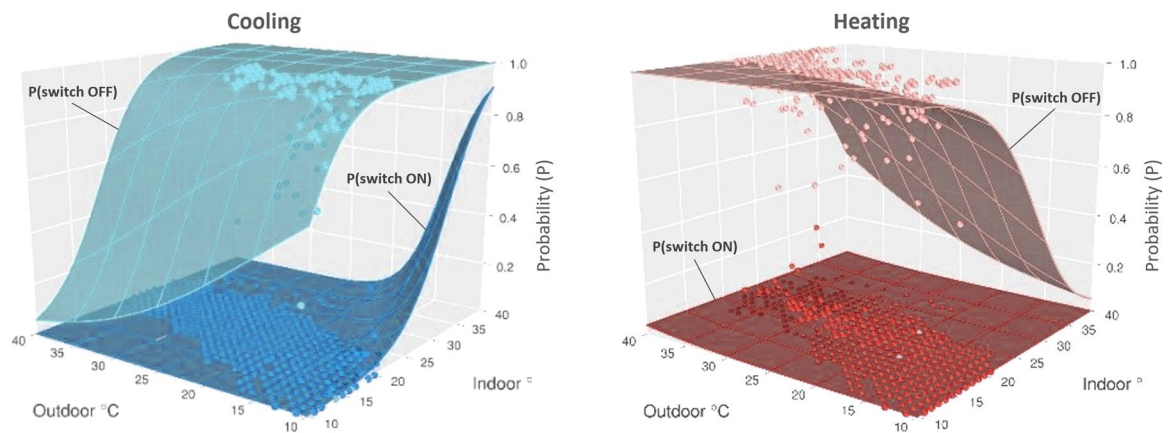

Fig. 12 An example for an application of the dataset: creating a model of occupant behaviour to predict the switching of air-conditioners, for both heating and cooling. This example model uses multiple logistic regression, with indoor and outdoor temperatures as the independent variables.

4. *HR.csv* contains the average heart rate data extracted from the BVP signals, calculated in spans of 10 seconds. The first row is the initial time of the session and it is 10 seconds after the beginning of the recording. The sampling rate of heart rate is 1 Hz.
5. *IBI.csv* contains the time intervals between a participant's heartbeats extracted from the BVP signals. This file does not have a sampling rate. The first column is the time (with respect to the starting time) of the detected inter-beat interval expressed in seconds (s). The second column is the duration in seconds (s) of the detected inter-beat interval (i.e., the distance in seconds from the previous beat).
6. *TEMP.csv* contains data from a temperature sensor expressed in degrees Celsius (°C), sampled at 4 Hz.

Class_wearable_data. The *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 of 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.

Technical Validation

Indoor and outdoor environmental data. Tables 5 and 6 show the specifications of the outdoor and indoor weather stations, respectively. Figure 8 shows a sample of data captured from these sensors and the air-conditioning states that we inferred from the data. Figure 9 shows data captured by the indoor weather station. Figure 10 shows data captured by the outdoor weather station. We were not able to find the accuracy of the Philio sensors in their datasheet, but assumed that it was sufficient for the purposes of our study. Figures 11 and 12 demonstrate a potential use case for the longitudinal dataset; here, we fitted different logistic regression models to the air-conditioning usage data. Within the scope of our research, we were not able to validate these models. The figures serve merely as an illustration of what the dataset may be used for.
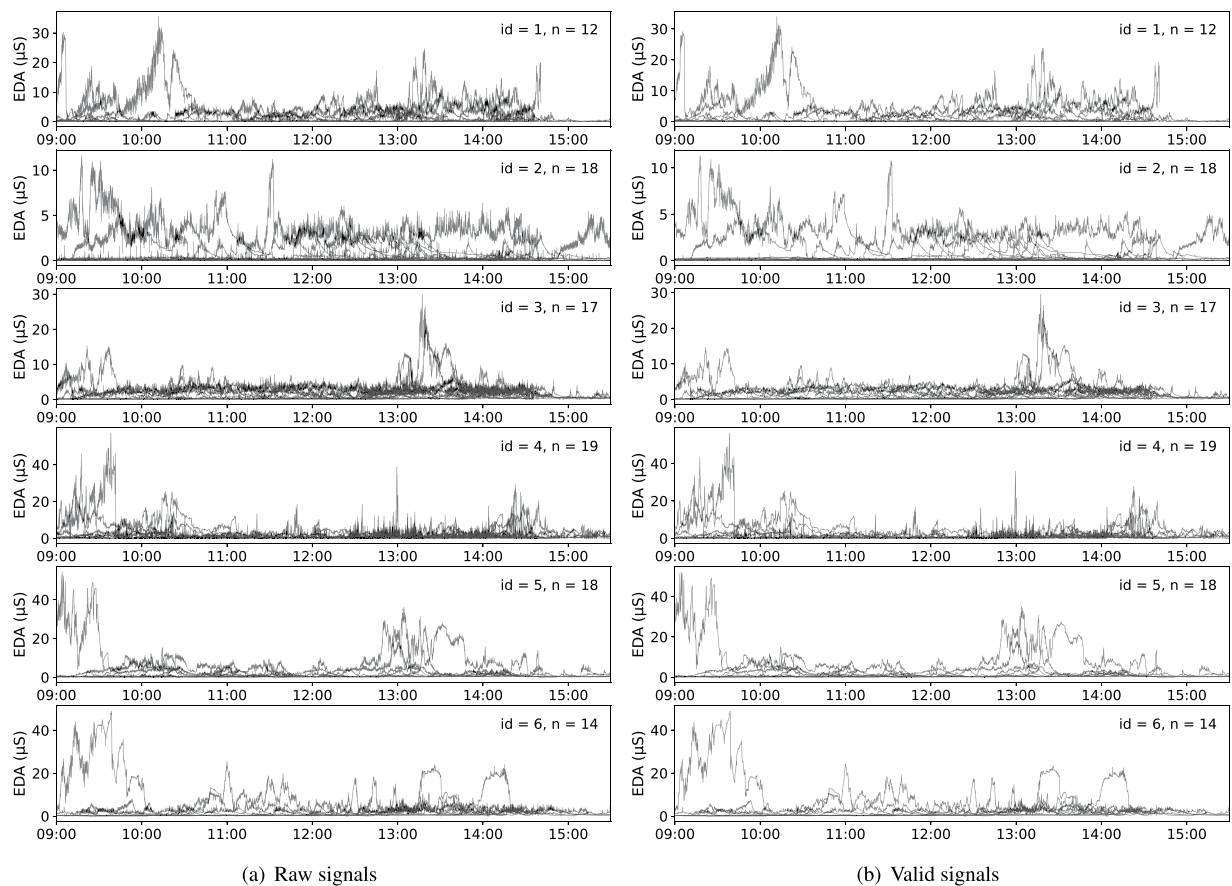
Fig. 13 EDA signals per school day for six participants before and after quality control.

Wearable signals. In the dataset, wearable signals include EDA, HRV, skin temperature and 3-axis acceleration. The above signals can be utilized to understand people's behaviours, engagement, thermal comfort and emotion. Before the data analysis and modelling, the data cleaning stage needs to be conducted to remove noises and motion artefacts for quality control[21]. Usually, there are several types of noises during the data collection from E4 wristbands[9]: flat responses, abrupt signal drops and quantization errors. Flat responses (i.e., 0 micro siemens) happen when there is poor contact between skin and the wristband, and abrupt signal drops occur with movements of wristbands. The quality of wearable measurements in the dataset has been thoroughly examined.

Missing values. Though physiological signals are mostly error-free for the majority of files in the dataset, a portion of data is missing due to issues of devices or human errors (e.g., poor contact, exhausted battery, unintentionally turning-off wristbands, unwilling to keep wearing wristbands). On average, participants contribute wearable signals for 15 days, with different wearing times during the school day (9:00–15:30). P13 and P16 contribute the least number of wearing days ($n = 6$) and P21 wear the wristbands for most days ($n = 28$). Then we calculate the missing values (i.e., 0 micro siemens for EDA signal) due to flat responses or motion artefacts. On average, 2.66% of data is missing for all participants, with P10 having the maximal portion of missing values (27.37%) and P20 having the least missing values (0.01%).

Quality control. Figure 13 shows the EDA signals in the school day for six participants before and after applying quality control. Due to space limitations, we only show the data from six participants where n indicates the number of wearing days of wristbands. The quality control involved the removal of motion artefacts (MAs). The median filter with a 5-second window is applied on EDA signals per school day as in previous research[9]. After quality control, the portion of missing values decreases from 2.66% to 2.46%. To better control the quality of wearable signals, various methods can be applied such as visual inspection[9], different filters with different settings, shape-based artefacts detection[22], etc. However, the aforementioned methods for different types of wearable signals (e.g., EDA, HRV) are usually selected or adapted according to the specific needs and purposes of researchers. Most importantly, unlike environmental data, there is no absolute ground truth for the physiological signals (EDA, HRV, and skin temperature) and physical activities (3-axis acceleration) of people in the real world. It does not help much to present a fixed and complete pipeline for the quality control of wearable signals. Therefore we only consider dealing with the missing values with a median filter for the EDA data validation as an example of quality control.

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[23]. 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[24], clustering[25,26]) and modelling techniques[27–29] could be explored to build prediction models for measuring occupants' mental state 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*[9,30]. 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[9,31]. 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*[32]. It could be helpful to explore group-wise seating behaviours and their relationship to perceived engagement and physiological synchrony[33]. (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[34], and teachers could ventilate classrooms timely to prevent excess carbon dioxide from affecting students' concentration[35–37].

Some limitations of the datasets need to be addressed. Firstly, Empatica E4 wristbands are susceptible to noises caused by many factors, e.g., loose electrodes and faulty wiring[38], and other devices such as indoor and outdoor weather stations may also be subject to similar systematic errors. Additionally, the accuracy of measuring physiological and behavioural data using E4 wristbands is limited. Menghini *et al.*[39] found that similar accuracy could not be achieved when comparing EDA signals between wrists and fingers. Lead wire extension is one promising solution to improve the accuracy of E4 wristbands, which allows EDA recordings to be moved from the wrist to the surface of the fingers or palm, thus eliminating the potential site differences. In our data collection, using the E4 wristband is the best option for collecting data from student and teacher participants on-site without putting an extra burden on them. Finally, besides the quality control for EDA measurements, additional measures could be further employed to validate the collected data (e.g., self-report data and other physiological signals) through appropriately designed studies.

Code availability

Python code for prepossessing the data and implementing the segmentation based on different classes are available online https://github.com/cruiseresearchgroup/InGauge-and-EnGage-Datasets.

References

1. Haldi, F. & Robinson, D. The impact of occupants' behaviour on building energy demand. *Journal of Building Performance Simulation* **4**, 323–338 (2011).
2. Rijal, H. B., Tuohy, P., Humphreys, M. A., Nicol, J. F. & Samuel, A. An algorithm to represent occupant use of windows and fans including situation-specific motivations and constraints. In *Building Simulation*, **4**, 117–134 (Springer, 2011).
3. Schiavon, S. & Lee, K. H. Dynamic predictive clothing insulation models based on outdoor air and indoor operative temperatures. *Building and Environment* **59**, 250–260 (2013).
4. Schweiker, M., Haldi, F., Shukuya, M. & Robinson, D. Verification of stochastic models of window opening behaviour for residential buildings. *Journal of Building Performance Simulation* **5**, 55–74 (2012).
5. Langevin, J., Gurian, P. L. & Wen, J. Tracking the human-building interaction: A longitudinal field study of occupant behavior in air-conditioned offices. *Journal of Environmental Psychology* **42**, 94–115 (2015).
6. Reinhart, C. F. & Voss, K. Monitoring manual control of electric lighting and blinds. *Lighting Research & Technology* **35**, 243–258 (2003).
7. Cheung, T., Schiavon, S., Parkinson, T., Li, P. & Brager, G. Analysis of the accuracy on pmv–ppd model using the ashrae global thermal comfort database ii. *Building and Environment* **153**, 205–217 (2019).
8. Kim, J. & de Dear, R. Thermal comfort expectations and adaptive behavioural characteristics of primary and secondary school students. *Building and Environment* **127**, 13–22 (2018).
9. Gao, N., Shao, W., Rahaman, M. S. & Salim, F. D. 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**, 1–26 (2020).
10. Di Lascio, E., Gashi, S. & Santini, S. Unobtrusive assessment of students' emotional engagement during lectures using electrodermal activity sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* **2**, 1–21 (2018).
11. Bakker, J., Pechenizkiy, M. & Sidorova, N. What's your current stress level? detection of stress patterns from gsr sensor data. In *2011 IEEE 11th International Conference on Data Mining Workshops*, 573–580 (IEEE, 2011).
12. Sarchiapone, M. *et al.* The association between electrodermal activity (eda), depression and suicidal behaviour: A systematic review and narrative synthesis. *BMC Sychiatry* **18**, 1–27 (2018).
13. Pollak, J. P., Adams, P. & Gay, G. Pam: a photographic affect meter for frequent, *in situ* measurement of affect. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 725–734 (2011).
14. Garbarino, M., Lai, M., Bender, D., Picard, R. W. & Tognetti, S. Empatica e3—a wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition. In *2014 4th International Conference on Wireless Mobile Communication and Healthcare-Transforming Healthcare Through Innovations in Mobile and Wireless Technologies*, 39–42 (IEEE, 2014).
15. Gao, N., Shao, W. & Salim, F. D. Predicting personality traits from physical activity intensity. *Computer* **52**, 47–56 (2019).
16. Michael, P. R., Johnston, D. E. & Moreno, W. A conversion guide: solar irradiance and lux illuminance. *Journal of Measurements in Engineering* **8**, 153–166 (2020).
17. Wagner, A., O'Brien, W. & Dong, B. *Exploring occupant behavior in buildings: methods and challenges* (Springer International Publishing, Cham, 2018).

18. Gao, N., Marschall, M., Burry, J., Watkins, S. & Salim, F. In-Gauge and En-Gage datasets. *Figshare* https://doi.org/10.25439/rmt.14578908 (2021).
19. Taylor, B. N. & Thompson, A. *The international system of units (SI)* (US Department of Commerce, Technology Administration, National Institute of Standards and Technology, 2001).
20. Braithwaite, J. J., Watson, D. G., Jones, R. & Rowe, M. A guide for analysing electrodermal activity (eda) & skin conductance responses (scrs) for psychological experiments. *Psychophysiology* **49**, 1017–1034 (2013).
21. Babaei, E., Tag, B., Dingler, T. & Velloso, E. A critique of electrodermal activity practices at chi. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–14 (2021).
22. Gashi, S. *et al.* Detection of artifacts in ambulatory electrodermal activity data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* **4**, 1–31 (2020).
23. Schwee, J. H. *et al.* Room-level occupant counts and environmental quality from heterogeneous sensing modalities in a smart building. *Scientific data* **6**, 1–11 (2019).
24. Deldari, S., Smith, D. V., Sadri, A. & Salim, F. Espresso: Entropy and shape aware time-series segmentation for processing heterogeneous sensor data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* **4**, 1–24 (2020).
25. Shao, W., Salim, F. D., Song, A. & Bouguettaya, A. Clustering big spatiotemporal-interval data. *IEEE Transactions on Big Data* **2**, 190–203 (2016).
26. Shao, W. *et al.* Onlineairtrajclus: An online aircraft trajectory clustering for tarmac situation awareness. In *2019 IEEE International Conference on Pervasive Computing and Communications* (*PerCom*, 192–201 (IEEE, 2019).
27. Salim, F. D. *et al.* Modelling urban-scale occupant behaviour, mobility, and energy in buildings: A survey. *Building and Environment* **183**, 106964 (2020).
28. Carlucci, S. *et al.* Modeling occupant behavior in buildings. *Building and Environment* **174**, 106768 (2020).
29. Kjærgaard, M. B. *et al.* Current practices and infrastructure for open data based research on occupant-centric design and operation of buildings. *Building and Environment* **177**, 106848 (2020).
30. Rahaman, M. S. *et al.* An ambient–physical system to infer concentration in open-plan workplace. *IEEE Internet of Things Journal* **7**, 11576–11586 (2020).
31. Gashi, S., Di Lascio, E. & Santini, S. Using students' physiological synchrony to quantify the classroom emotional climate. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, 698–701 (2018).
32. Sacerdote, B. Experimental and quasi-experimental analysis of peer effects: two steps forward? *Annu. Rev. Econ.* **6**, 253–272 (2014).
33. Gao, N., Rahaman, M. S., Shao, W. & Salim, F. D. 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*, 237–242 (2021).
34. Park, R. J., Goodman, J., Hurwitz, M. & Smith, J. Heat and learning. *American Economic Journal: Economic Policy* **12**, 306–39 (2020).
35. Arief-Ang, I. B., Salim, F. D. & Hamilton, M. Da-hoc: semi-supervised domain adaptation for room occupancy prediction using co2 sensor data. In *Proceedings of the 4th ACM International Conference on Systems for Energy-Efficient Built Environments*, 1–10 (2017).
36. Arief-Ang, I. B., Hamilton, M. & Salim, F. D. Rup: Large room utilisation prediction with carbon dioxide sensor. *Pervasive and Mobile Computing* **46**, 49–72 (2018).
37. Arief-Ang, I. B., Hamilton, M. & Salim, F. D. A scalable room occupancy prediction with transferable time series decomposition of co2 sensor data. *ACM Transactions on Sensor Networks (TOSN)* **14**, 1–28 (2018).
38. McCarthy, C., Pradhan, N., Redpath, C. & Adler, A. Validation of the empatica e4 wristband. In *2016 IEEE EMBS international student conference (ISC)*, 1–4 (IEEE, 2016).
39. Menghini, L. *et al.* Stressing the accuracy: Wrist-worn wearable sensor validation over different conditions. *Psychophysiology* **56**, e13441 (2019).
40. Healey, J. A. & Picard, R. W. Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on Intelligent Transportation Systems* **6**, 156–166 (2005).
41. Koelstra, S. *et al.* Deap: A database for emotion analysis using physiological signals. *IEEE Transactions on Affective Computing* **3**, 18–31 (2011).
42. Schneegass, S., Pfleging, B., Broy, N., Heinrich, F. & Schmidt, A. A data set of real world driving to assess driver workload. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 150–157 (2013).
43. Wang, R. *et al.* Studentlife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 3–14 (2014).
44. Abadi, M. K. *et al.* Decaf: Meg-based multimodal database for decoding affective physiological responses. *IEEE Transactions on Affective Computing* **6**, 209–222 (2015).
45. Birjandtalab, J., Cogan, D., Pouyan, M. B. & Nourani, M. A non-eeg biosignals dataset for assessment and visualization of neurological status. In *2016 IEEE International Workshop on Signal Processing Systems (SiPS)*, 110–114 (IEEE, 2016).
46. Subramanian, R. *et al.* Ascertain: Emotion and personality recognition using commercial sensors. *IEEE Transactions on Affective Computing* **9**, 147–160 (2016).
47. Gjoreski, M., Luštrek, M., Gams, M. & Gjoreski, H. Monitoring stress with a wrist device using context. *Journal of Biomedical Informatics* **73**, 159–170 (2017).
48. Schmidt, P., Reiss, A., Duerichen, R., Marberger, C. & Van Laerhoven, K. Introducing wesad, a multimodal dataset for wearable stress and affect detection. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, 400–408 (2018).
49. Gjoreski, M. *et al.* Datasets for cognitive load inference using wearable sensors and psychological traits. *Applied Sciences* **10**, 3843 (2020).
50. Park, C. Y. *et al.* K-emocon, a multimodal sensor dataset for continuous emotion recognition in naturalistic conversations. *Scientific Data* **7**, 1–16 (2020).
51. Handbook-Fundamentals, A. American society of heating. *Refrigerating and Air-Conditioning Engineers* (2009).
52. Fuller, K. A. *et al.* Development of a self-report instrument for measuring in-class student engagement reveals that pretending to engage is a significant unrecognized problem. *PloS One* **13**, e0205828 (2018).

Acknowledgements

Author contributions

N.G. designed, prepared, and conducted the cross-sectional study for wearable data collection, analysed the wearable and indoor sensor data, and wrote the manuscript. M.M. designed, prepared and conducted the longitudinal data collection for outdoor and indoor sensing, and processed the dataset and wrote the manuscript. J.B., S.W. and F.D.S. supervised the data collection, dataset design, and revised the manuscript. F.D.S. advised N.G. on the overall project and data analysis and modelling.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to F.D.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.