Building and Environment 203 (2021) 108037

Contents lists available at ScienceDirect

Building and Environment

journal homepage: www.elsevier.com/locate/buildenv

Meta-analysis of 35 studies examining the effect of indoor temperature on office work performance

Jose Ali Porras-Salazar [a,*], Stefano Schiavon [b], Pawel Wargocki [c], Toby Cheung [a], Kwok Wai Tham [d]

[a] Berkeley Education Alliance for Research in Singapore, Singapore
[b] Center for the Built Environment, University of California, Berkeley, CA, USA
[c] International Centre for Indoor Environment and Energy, Technical University of Denmark, Kongens Lyngby, Denmark
[d] Department of Building, National University of Singapore, Singapore

ARTICLE INFO

Keywords:
Indoor air temperature
Thermal environment
Offices
Productivity
Cognitive performance

ABSTRACT

Several relationships between air temperature and work performance have been published. We reanalysed the one developed in 2006 by Seppänen et al.; which is probably the best known. We found that even when significant, its prediction accuracy is very low ($R^2 = 0.05$, $MAE = 1.9\%$, $RMSE = 3.1\%$). We consequently reviewed the literature and found 35 studies on the effects of temperature on office work performance. We used Seppänen et al.'s approach to normalise the data reported in these studies and explored the feasibility to develop a new relationship using regression models, models based on the Maximal Adaptability framework, and machine learning. We could not find a relationship between temperature and office work performance neither for the range of temperatures measured in most of the office buildings (20 °C–30 °C) or a wider range (18 °C–34 °C). Plausible reasons are discussed including the variety of methods used to assess performance, the multiple uncontrolled confounders, and the fact that temperature alone may not fully describe how the thermal environment affects building occupants. We do not recommend the use in practice of any of the models relating temperature to office work performance examined in the present study. The lack of relationships does not necessarily refute that temperature affects the performance of office work. Coordinated research predicated on a shared protocol enabling integrated analysis in the modelling of the relationships between the indoor thermal environment and office work performance is proposed to be carried out before using them in practice. We made the database open-source and developed an application for data exploring.

1. Introduction

Indoor environmental conditions affect physiological reactions and psychological responses, which in turn can affect the psychomotor, perceptual, and cognitive abilities needed to perform work [1]. Given that in high-income countries the greatest proportion of operating expenditures are the salaries of office workers [2–5], changes in work performance have strong socio-economic implications [6,7]. The thermal environment is one component of indoor environmental quality and its conditions in the workspace can affect employee performance. This was documented in a wide body of publications, whereby objective (e.g., simulation tasks, psychological tests) [8,9], subjective (e.g., self-reported performance, peer-evaluations) [10,11], and physiological (e.g., EEG, heart rate variability) [12–14] methods were used to evaluate work performance.

To examine and demonstrate the potential economic benefits of improving indoor thermal conditions, several studies summarised in Table 1 created relationships between the thermal environment (i.e., indoor air temperature or thermal response) and performance [4,9,12,15–23]. However, these efforts have some limitations that make it difficult to use them to estimate the size of the effect on work performance that spans the holistic set of psychomotor, perceptual, and cognitive skills. Some studies merged data from experiments carried out with different populations (e.g., children and adults, students and workers, office workers and industrial workers) [18,20], others only used the results obtained in their measurement campaigns [9,12,22], or used a temperature range far beyond the normal thermal conditions in office buildings [18,23], or favoured the use of some performance

* Corresponding author.
E-mail address: jose.porrassalazar@ucr.ac.cr (J.A. Porras-Salazar).

https://doi.org/10.1016/j.buildenv.2021.108037
Received 22 March 2021; Received in revised form 2 June 2021; Accepted 4 June 2021
Available online 11 June 2021

Table 1
Previously developed relationships describing how work performance is affected by the indoor thermal environment.

Author		Year	Proxy of the thermal environment	Description
Ramsey and Morrisey	[15]	1978	WGBT	Used 102 measures of performance from 22 studies. Developed a series of iso-decrement curves showing how performance changes as a function of WGBT and time. Focussed on hot environments only.
Wyon	[16]	1986	Air temperature	NA
Berglund et al.	[17]	1990	ET*	NA
Roelofsen	[4]	2002	PMV	Used data from two studies. Related the loss of performance with PMV
Pilcher et al.	[18]	2002	WBGT	Used data from 22 studies (515 measures of performance) and coded them into five categories: Cold 2, Cold 1, Hot 1, Hot 2, and Hot 3, and a mean percent difference in performance was estimated between the neutral temperature group and each of the five temperature subcategories. Performance decreased by ~14% when the WBGT exceeds 32.2 °C or was below 10.8 °C (DBT [a] 37 °C and 14 °C). Temperatures (DBT) between 25 °C and 31 °C had little effect on performance.
Kosonen and Tan	[19]	2004	PMV Air temperature	Used data from three studies. A peak level of productivity was achieved when the temperature was 20 °C and PMV is −0.21
Seppänen et al.	[20]	2006	Air temperature	Used data from 24 studies and ~130 data points [b]. The measures of performance were normalised using the percentage change in performance per 1 °C increase in temperature. Normalised observations were regressed against the average temperature. Maximum performance was achieved around 22 °C.
Jensen et al.	[21]	2009	TSV	Used data from three studies (339 TSVs) performed in mechanically ventilated buildings. Individual measures of performance were regressed against the TSV. Optimal performance was achieved when the thermal environment was perceived as slightly cool.
Lan et al.	[12]	2011	Air temperature TSV	Used data from their study and two other studies carried out by the main author (57 participants and 2692 TSVs). Individual measures of performance were regressed against the temperature and TSV. In the former optimum performance is achieved at 22 °C, while in the latter when people felt between neutral and slightly cool.
Cui et al.	[9]	2013	Air temperature TSV	Used data from their study (36 participants). The mean memory typing performance at 22 °C, 24 °C, 26 °C, 29 °C, and 32 °C, was regressed against the temperature and TSV. Maximum performance was achieved at 0.14 TSV.
Geng et al. [c]	[22]	2017	Air temperature TSV Thermal Satisfaction Votes	Used data from their study (21 participants). The mean performance at temperatures between 16 °C and 28 °C was regressed against the temperature, TSV and thermal satisfaction votes. Maximum performance was achieved around 23 °C and when participants felt neutral.
Yeganeh et al.	[23]	2018	Air temperature	Used data from 28 studies (481 measures of performance). The measures of performance were grouped into five categories: Cold 2, Cold 1, Hot 1, Hot 2, and Hot 3, and the percentage difference in performance between adjacent groups was estimated based on the effect sizes. The relationship presents a bell-shaped curve where 100% cognitive performance was achieved around 22 °C.

WBGT: Wet-bulb globe temperature index.
ET*: New effective temperature.
PMV: Predicted mean vote.
TSV: Thermal sensation votes.
[a] Assuming that relative humidity is 60% and that dry-bulb and globe temperatures are equal.
[b] Seppänen et al. [20] included two studies in classrooms with 9 to 11-year-old children and two studies in universities.
[c] Geng et al. [22] and Cui et al. [9] were the only ones who provided measures of the accuracy of the models. Geng et al. reported an R^2 of ≈0.9 for the temperature and TSV relationships and 0.75 for the thermal satisfaction model. Cui et al. reported an R^2 of 0.99 for the temperature relationship and of 0.97 for the TSV model.

measures over others [9,12,22]. The most important limitation of the proposed models was non-reporting prediction power and model accuracy [4,12,15–23].

Among the relationships shown in Table 1, the best known and most widely used relationship between temperature and work performance in the field is that published by Seppänen et al., in 2006 (Fig. 1b) ([20]: 391 citations, [24]: 120 citations, [25]: 252 citations). It has been reported in the REHVA Guidebook 6 and used to integrate work performance in the life-cycle cost analysis of building services [5]; publications showing and estimating economic impacts of heat on labour and learning [26–29]; studies estimating the economic benefits of improving indoor thermal conditions [30–32], and the ASHRAE Handbook- Fundamentals [33] where it is the basis of the relationship between office work performance and the deviation from optimal comfort temperature.

Seppänen et al. [20] used 24 studies published between 1970 and 2004 to develop their relationship; eight of them were performed in offices, nine in laboratories, three in field laboratories, and four in classrooms. Two of the studies in classrooms were with children aged between 9 and 11-years, and two other studies were with university students. While we estimated that their relationship was based on approximately 130 data points (see section 2.2.), an updated version consisting of 148 data points from 26 studies had been published by Seppänen and Fisk [25] in the same year. The data were normalised using the change in performance per 1 °C increase in temperature (λ). Then, the λ values were regressed against the average temperature —estimated based on the experimental range of temperature conditions—, and a quadratic polynomial model was fit to the data points (Fig. 1a). The equation representing this polynomial was not provided. To make the graph more communicative, Seppänen et al. [20] transformed this first model into a second one, quantifying the relationship between Relative Performance (*RP*) and temperature (Fig. 1b). The relationship presented in Equation (1) indicates the maximum performance at 21.8 °C while the work performance will decrease below 21–22 °C and above 23–24 °C. From 15 °C to 35 °C, a reduction in performance of about 1% is expected for every ±1 °C change of temperature from 22 °C when using this relationship. According to the relevance of the metric on overall work performance and sample size, different weighting factors were applied to the performance outcomes.

$$RP = -0.4685328 + 0.1647524 \cdot T - 0.0058274 \cdot T^2 + 0.0000623 \cdot T^3 \quad (1)$$

Where *RP* is performance relative to the maximum value and *T* is room temperature in °C.

Although Seppänen et al. [20] wrote in their publication that the "This relationship has a high level of uncertainty; however, use of this relationship may be preferable to current practice, which ignores productivity […]", neither the statistical measures of the uncertainty nor the model's statistical significance were provided. The high level of uncertainty indicated by the authors could be due to the diverse and apparent contradictory findings shown within the studies included in

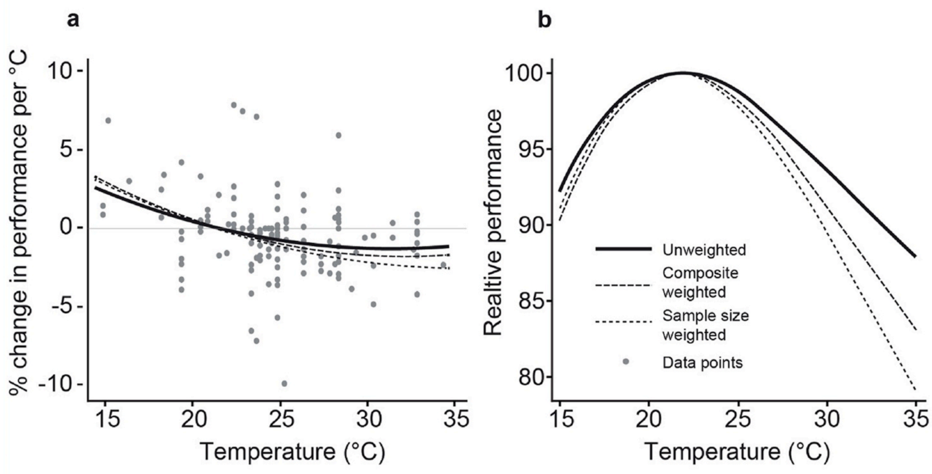

Fig. 1. Replication of Seppänen et al.'s model [20] based on data extraction. Fig. 1a Relationship between the percentage change in performance per 1 °C increment and temperature. Fig. 1b Relationship between temperature and relative performance. For visualization efficiency, one data point was removed from Fig. 1a (but not from the analysis) because was above 10%.

their database. This has been pointed out by other researchers when proposing their relationships and is attributed to methodological reasons and potential confounding factors [15,23,34,35]. In addition to the indoor thermal predictor, there could also be other factors influencing work performance. Zhang et al. [34] classified these factors as those that are environment-related (e.g., the intensity of thermal stress and duration of exposure), task-related (e.g., task type and task complexity), and performer-related (e.g., skill level, gender and age, and thermal acclimatisation). Task complexity, speed and accuracy, and climate, which have been broadly described in the literature, are highlighted below.

1.1. Task complexity

Task complexity seems to be one of the main reasons for the diverse and inconsistent findings [36]. It is generally accepted that simpler cognitive tasks are less sensitive to heat stress than complex ones [18,36,37]. Therefore, several authors [15,18,23,36,38,39] have tried to categorise cognitive tasks according to their complexity based on the mental effort required to perform them. Taylor et al. [39] classified the most common tasks as "simple" or "complex". While the former requires very simple perceptual-motor skills to complete (e.g., choice reaction time, memory recall), the latter requires more effort and/or attention, such as multiple or dual tasks (e.g., complex motor coordination and working memory tasks). Hancock et al. [35] proposed three subdivisions (i.e., psychomotor, perceptual, and cognitive response capacities) based on the differentiation of information-processing stages. Yeganeh et al. [23] grouped their tasks based on both Hancock's subdivision and the simple-complex categorisation method, mentioning that simple tasks are those psychomotor and perceptual-related, while complex tasks are the ones that are cognition-related. As brain imaging techniques have become more widely used in studies examining the effects of the indoor environment on work and learning performance, some results [13,40] have shown that tasks activate different areas of the brain with different levels of intensity. Therefore, denoting tasks as "simple" or "complex" —based only on expert judgment— may be too simplistic [39].

1.2. Speed and accuracy

Speed and accuracy are two of the most used metrics characterising manual, perceptual, and cognitive performance [41]. Speed measures how quickly participants perform the work task, irrespective of whether the answer is correct or not. Response time can also be considered as a measure of speed since it quantifies how quickly a participant reacts to a stimulus. Frequently used metrics of speed considered are the number of exercises completed in a certain period, reaction time, movement time, and latency. Accuracy is measuring how correct the task is completed, and is often measured by the number of correct exercises, scores, errors, or false positives, etc. Performance can also be measured by creativity; however, this metric was uncommonly used and we decided not to discuss it in detail here. The probability of making a correct decision increases as information accumulates, but more time is required in the gathering process [42]. Therefore, faster responses involve less accumulated evidence and hence less informed decisions with a higher probability of being wrong [42,43]. People tend to accept an error rate if it means saving time and resources. This momentary desire to trade speed for accuracy —or vice versa— is described by researchers as the speed-accuracy trade-off (SAT) [43,44].

1.3. Type of climate

Since the type of climate may have an effect on acclimatisation and adaptation to the thermal environment, this influences the relationship between the thermal environment and work performance. Although people living in hot environments (e.g., tropical and hot dry climates) would have a higher tolerance to heat due to acclimatisation [1], this is not always the case. For example, while people that exercise regularly in temperate countries may be acclimatised to the heat as well, people who spend all their time in the air-conditioned spaces in warm regions do not [45]. In previous literature reviews, only Wargocki et al. [46] separated the studies performed in the tropical climate from the ones in temperate and continental climates. However, their meta-analysis was based on studies where all participants were children.

1.4. Objectives

We integrated all identified data on the effects of the thermal environment on performance to create relationships alternative to the one developed by Seppänen et al. [20]. We did this for two reasons: (i) since the development of the work by Seppänen et al. [20] in 2006, several studies have been published allowing their data to be compiled into a single database and reverified the relationship between temperature and work performance, and (ii) having an updated relationship would help the building industry stakeholders estimate the potential economic benefits of improving the indoor thermal conditions. Instead of including work performance measurements across different spaces (e.g., schools as done by Seppänen et al. [20]), we focussed on office work environments and reviewed studies on the effect of temperature on cognitive abilities essential for office work (e.g., quantitative knowledge or short-term memory) or office work itself. As a result, the objectives of our work were:

1. To summarise all selected publications on the effects of thermal environment on office work into an open-source database through a meta-analysis;
2. To provide the measures of prediction accuracy for the relationship proposed by Seppänen et al. [20] using the original data and the data from our database;
3. To propose alternative relationships between work performance and temperature using our database;
4. To analyse the effects of potential confounders in the developed relationships, i.e., task complexity, task speed/accuracy, and climate.

2. Methods

2.1. Developing a database

2.1.1. Search strategies and data collection

We searched electronic databases containing scientific publications from September 2019 to March 2020, which included: Google Scholar, Web of Science, Elsevier, PubMed, and ProQuest. A variety of keyword combinations were used (See Appendix A in the Supplementary Material (SM)). We also reviewed the reference section of Seppänen et al. [20], Ramsey and Morrissey [15], Pilcher et al. [18], Hancock et al. [35], Yeganeh et al. [23], and Zhang et al. [34].

We only searched for peer-reviewed journal articles that reported measurements of both the thermal environment and performance of office work. The measurements of the thermal environment included objective measures of temperature or operative temperature, but not subjective ratings of thermal sensation or thermal comfort. Diverse measures were considered to be relevant for office work performance, which comprised of diagnostic tests, simulated office work tasks, and existing outcome metrics.

Studies that used only self-estimated performance (e.g., Tanabe et al. [47] Lamb and Kwok [11]) were not included, neither were proxies for reduced performance, such as the prevalence and intensity of neurobehavioral symptoms, especially fatigue, difficulty concentrating, sleepiness, or headaches. Since the above measures did not provide an objective measure of performance, they were not included. Although offering analyses of long-term effects, absenteeism and presentism can be affected by many uncontrollable factors prone to spurious conclusions thus speculative in nature. For these reasons, results predicated only on worker absenteeism and presentism were also not considered. Physiological measurements providing information on cognitive load (e.g., EEG, ECG, heart rate variability, pupillary responses) have not been considered. Because aspects of the built environment and other parameters that may affect performance differ substantially between industrial work, learning, and office work [48], data from factory workers, university students, and primary/secondary school children were also excluded.

We characterised the thermal environment by air temperature in our research scope since it was the most frequently reported parameter in the identified studies. When the operative temperature was reported instead of air temperature, we assumed equivalence between ambient and radiant temperature (i.e., the operative temperature was directly used without any conversion [49]). We excluded studies where participants were exposed to extreme thermal conditions (e.g., Refs. [50,51]). To be specific, we did not include any measures of performance where the range of experimental temperatures was below 17 °C or above 36 °C. The temperature range we used in this study matches ~99% of the office data found in the ASHRAE Global Thermal Comfort Database II (66251 data points) [52]. Finally, we excluded all studies where thermal stress was induced by any means other than the indoor thermal conditions (e.g., exercise or water immersion).

From each study, we retrieved the year of publication, journal, and information regarding the study location, whether the study was or was not performed in a controlled environment, the sample size, age group, occupation of the participants, clothing, and physical activity level. We also collected the tasks or tests used to measure performance, the performance metrics and outcomes, and the temperature conditions to which the participants were exposed. Table 2 summarises the collected information, and the database with detailed information can be found in DRYAD [53]. We developed an application for the interactive visualization of the data (http://cbe-berkeley.shinyapps.io/temp-performance). All reported measures of performance were included in our database, regardless of the level of statistical significance.

2.1.2. Data normalisation

For most studies in our database, the results do not report the standard deviation or effect size. Performance data available in the literature is mainly reported in two ways: absolute data (i.e., number of completed exercises) and percentage data (e.g., percentage of errors) at a particular temperature set-point. We normalised the performance data obtained from the studies using the method proposed by Seppänen et al. [20,54], which was similarly used by Wargocki et al. [46,55]. This approach assumes that performance changes linearly within the high temperature (T_H) and low temperature (T_L) range examined in each study regardless of the performance measure used and the temperature range. Equation (2) calculates the change in work performance in % per 1 °C increment in temperature ($\lambda_{\%}$), whereby positive $\lambda_{\%}$ indicates an increase in performance with increasing temperature; while negative $\lambda_{\%}$ indicates a decrease in performance with increasing temperature.

$$\lambda_{\%} = \frac{P(T_H) - P(T_L)}{P(T_L)} \cdot 100 \cdot \frac{1}{T_H - T_L} \tag{2}$$

$P(T_H)$ and $P(T_L)$ respectively represents the mean task performance index for participants at the high- and low-temperature conditions, whereby the input value is dependent on how the results of performance were reported in the studies (i.e., absolute data or percentage data. See Table B1 in SM). An example of an absolute outcome is a subject that accurately answers nine questions at 25 °C and seven at 30 °C ($P(T_H) = 7$ and $P(T_L) = 9$); while for percentage outcome, this is a subject that accurately answers 90% of the questions at 25 °C and 70% at 30 °C ($P(T_H) = 0.7$ and $P(T_L) = 0.9$). The calculated $\lambda_{\%}$ will be the same for both absolute and percentage data obtained from the literature. Depending on the specific performance metric used, an increase of the value could indicate either performance improvement or reduction. To ensure consistent results, sign control was applied and presented in Table B1 in the SM. Consequently, for those metrics (e.g., mean reaction time or the number of errors) where higher results would mean lower performance, the percentage change in performance ($\lambda_{\%}$) was multiplied by −1.

2.2. Assessing the performance of the model developed by Seppänen et al.

Despite the common usage of Seppänen et al.'s model [20], certain information (i.e., model equation and statistical evidence) was not provided. We decided to replicate their analysis by fitting the model to (i) the original data points to estimate its performance, and (ii) the data in our new database to validate its performance. In our analysis, we used the relationship published in Ref. [20] and not the one in Ref. [25] because the data points were easier to read and digitalise (i.e., more information was given regarding the procedure on how these data points were estimated). There are also some differences in the data presented in the figures found in both publications. In Seppänen and Fisk [25] there were 26 studies reported and the point where the slope equals zero was 21.6 °C, while there were only 24 studies mentioned in Ref. [20] and this value was 21.8 °C. The 2017 ASHRAE Handbook-Fundamentals [33] includes results from both publications.

We verified the first part of the model (i.e., (i)) showing the relationship between percentage change in performance per 1 °C increase in

Table 2
Summary of papers meeting the inclusion criteria: 29 papers reporting 35 studies.

Author		Year	Source	Location	Number of participants	Duration of exposure per temperature condition [d]
Mackworth [b]	[64]	1946	British Journal of Industrial Medicine	NA	11	180
Teichner and Wehrkamp [b]	[65]	1954	Journal of Experimental Psychology	NA	30	NA [e]
Chiles [b,c]	[60]	1958	Ergonomics	USA	11	25
					10	25
Pepler [b,c]	[63]	1958	Ergonomics	Singapore	16	160
					32	99
					8	154
					24	120
Givoni and Rim [a,b]	[66]	1962	Ergonomics	Israel	4	120
Griffiths and Boyce [b]	[67]	1971	Ergonomics	United Kingdom	50	NA [e]
Allen and Fisher [b]	[68]	1978	Ergonomics	USA	54	NA [e]
Bell [b]	[69]	1978	Human Factors	USA	144	30
Sharma and Panwar [b]	[70]	1987	International Journal of Biometeorology	India	53	180
Niemelä et al. [a,c]	[61]	2002	Energy and Buildings	Finland	35	NA
					15	NA
Vasmatzidis et al. [b]	[71]	2002	Ergonomics	USA	12	120
Tanabe and Nishihara [a]	[10]	2004	Indoor Air	Japan	40	90
Tham [a]	[8]	2004	Indoor Air	Singapore	56	80 h [f]
Witterseh et al. [a b]	[72]	2004	Indoor Air	Denmark	30	180
Tham and Willem [a (c)]	[62]	2005	ASHRAE Transactions	Singapore	27	80 h [f]
					26	80 h [f]
Lan and Lian [a b]	[73]	2009	Building and Environment	China	22	120
Lan et al. [a b]	[74]	2009	Building and Environment	China	24	80
Tham and Willem [a b]	[75]	2010	Building and Environment	Singapore	96	240
Lan et al. [a b]	[12]	2011	Energy and Buildings	China	12	270
Cui et al. [b]	[9]	2013	Building and Environment	China	36	150
Cui et al. [b]	[76]	2013	Building and Environment	China	18	150
Maula et al. [a b]	[77]	2015	Indoor Air	Finland	33	210
Trezza et al. [a b]	[78]	2015	AGE	Brazil	68	180
Geng et al. [a b]	[22]	2017	Building and Environment	China	21	120
Liu et al. [a b]	[79]	2017	Building and Environment	Denmark	12	180
Schiavon et al. [a]	[80]	2017	Indoor Air	Singapore	56	90
Zhang et al. [b]	[81]	2017	Building and Environment	Australia	26	180
Nayak et al. [a]	[13]	2018	Brain sciences	USA	7	155
Wang et al. [a b]	[14]	2019	Building and Environment	USA	15	120

[a] Metrics of speed were used.
[b] Metrics of accuracy were used.
[c] Publications that included more than one study.
[d] Time in minutes except when indicated.
[e] Exposure time < 3 h.
[f] Estimated based on the following assumption: 8 working hours per day and 5 days per week during 2 weeks.

temperature ($\lambda_{\%}$) against the average temperature (see Fig. 1a). Hence, we produced the equation using two different sets of data the *line dataset* and the *point dataset*. In the *line data set*, using Plot Digitizer [56], we digitalised 16 points along the model line (i.e., solid line without weighting in Fig. 1a) by measuring the percentage change in the performance value (y-axis) for every 2 °C in the graph. An additional point where the model line reaches y = 0 (21.8 °C) was also included. In the *point data set*, we digitalised a total of 130 data points corresponding to all reported measures of performance presented as dots in Fig. 1a. In both datasets, we fit a quadratic polynomial model. The equations of both models were compared and the correlation between them was estimated. A high correlation means the same result was achieved for the two different datasets; therefore, resulting in a higher probability that the equation we propose is similar to the one proposed by Seppänen et al. [20].

We used the developed model to estimate how well it fits to: (i) the original data points (*point dataset*) and (ii) the normalised data points from our database ($N = 358$). The prediction accuracy was determined by the coefficient of determination (R^2), the Mean Absolute Error (*MAE*), and the Root Mean Square Error (*RMSE*). R^2 shows the proportion of variance explained by the prediction, whereas *MAE* and *RMSE* measure the average magnitude of the error for a set of predictions. Since Seppänen et al.'s model [20] predicts λ_{mid} and not $\lambda_{\%}$ —as performed in Equation (2)—, before utilising this data, we first converted $\lambda_{\%}$ to decimals and used Equation (3) to transform our results into λ_{mid} values. According to Seppänen et al., λ_{mid} is the point estimate of λ at the midrange of temperatures [20]. The relationship between λ and λ_{mid} for different temperature differences can be found in Appendix C in the SM.

$$\lambda_{mid} = \frac{\lambda}{(1 + 0.5\,\lambda \cdot (T_H - T_L))} \tag{3}$$

2.3. Model development criteria

To update the relationship between work performance and temperature using our database, we followed Seppänen et al.'s approach [20] by regressing our $\lambda_{\%}$ values against the average temperatures (T_{avg}) (Equation (4)).

$$T_{avg} = \frac{T_H + T_L}{2} \tag{4}$$

According to the ASHRAE Global Thermal Comfort Database II (66251 data points) [52], 90% of the office buildings had an indoor temperature range between 20 °C and 30 °C. Since data beyond the normal temperature range in office buildings were captured in our database, we decided to perform the model selection analysis by excluding those measures of performance conducted under a temperature below 20 °C or above 30 °C ($T_L < 20$ °C and $T_H > 30$ °C).

2.3.1. Predictive model performance evaluation

We estimated the effects of indoor temperature on office work performance based on different prediction model algorithms (Table 3), selecting the models with the highest prediction accuracy, which was determined by the R^2, *MAE*, and the *RMSE*. In our approach, the data were separated into training and testing datasets using a 9:1 ratio and this was repeated 1000 times [57]. The performance indexes (R^2, *MAE*, and *RMSE*) were averaged among the 1000 repetitions in each model. All statistical analyses were performed using R software [58].

2.3.2. Confidence range

The best performing models from the previous analysis were brought into the next stage to develop a confidence range. Instead of using a single model to predict the values of $\lambda_{\%}$ and estimate its confidence limits, as Seppänen et al. [20] did, we repeated the simulations in each model 1000 times by randomly sampling 90% of the data for training and the remaining 10% for validation. For example, if the quadratic polynomial approach was selected, then 1000 quadratic polynomial models with different combinations of training and validating sample sets were created. These 1000 models were also used to calculate the median, the 2.5th and 97.5th percentiles at every temperature level (See Appendix F for details). Predictions were done with a step-size of 1 °C in the range of temperatures between 20 °C and 30 °C.

2.3.3. Model transformation

To help interpret the curves, we transformed them into a graph that quantifies the relationship between relative performance (*RP*) and temperature. Since the 2.5th, 50th, and 97.5th percentiles curves are not described by any equation, we fit a polynomial model onto the values of $\lambda_{\%}$. Later, we used these models and equation (5), which is an adapted version of the one presented by Seppänen et al. [54] to evaluate the ratio of performance between any two temperatures. When temperature increases from $T_{avg\,0}$ to $T_{avg\,1}$, the ratio of performance at $T_{avg\,1}$ ($P(T_{avg\,1})$) to that at $T_{avg\,0}$ ($P(T_{avg\,0})$) will be:

$$\frac{P(T_{avg1})}{P(T_{avg0})} = 100 \cdot \exp\left[1 \cdot \frac{1}{100} \cdot \int_{T_{avg0}}^{T_{avg1}} \widehat{\lambda}(T_{avg})\, dT_{avg}\right] \tag{5}$$

$\widehat{\lambda}(T_{avg})$ is the fractional change in performance per 1 °C at the temperature T_{avg} estimated by the fit. A step-by-step example of these calculations is given in Appendix F.

Table 3
Predictive model performance evaluation (20 °C–30 °C, $N_1 = 215$).

Model	Performance metrics		
	R^2	*MAE*	*RMSE*
Zero-order model [a]	0	1.6%	2.5%
Linear	0.05	1.6%	2.5%
Quadratic polynomial	0.07	1.6%	2.5%
Cubic polynomial	0.06	1.6%	2.5%
Quartic polynomial	0.07	1.6%	2.5%
Piecewise linear [b]	0.07	1.6%	2.5%
Generalised Linear Model (GLM)	0.05	1.6%	2.4%
k -Nearest Neighbours (kNN)	0.07	1.7%	2.6%
Support Vector Machines (SVM)	0.06	1.6%	2.5%
Random Forest (RF)	0.07	1.8%	2.8%
Locally Estimated Scatterplot Smoothing (LOESS)	0.06	1.6%	2.5%
Maximal Adaptability Model (MAM) [c d]	0.0	1.6%	2.6%

[a] Equation: $\lambda_{\%} = -0.334$.
[b] Breaking points at 23 °C, 25 °C, and 27 °C.
[c] Breaking points at 23 °C and 27 °C.
[d] For the *MAM*, we fit a single model and the R^2, *MAE*, and **RMSE** correspond to the performance metrics of that single model (See Appendix I).

Finally, the temperature(s) associated with the highest *RP* in the 50th percentile curve was considered as the condition under which office workers will achieve their maximum performance. Hence, we assigned it the value of 100%, whereby the performance at all other temperatures should be lower.

2.4. Multiple regression analysis

To analyse the effects of potential confounders (i.e., task complexity, the speed and accuracy metrics; and climate), we performed a multiple linear regression analysis where we dummy coded these categorical variables. For each one of the factors, we created *k*-1 dummy variables, where *k* is the number of categories in each variable, and the category that presented the highest number of data points was used in the analysis as the reference group (See Appendix G in the SM). We did not include the duration of exposure as a potential confounder in the multiple regression analysis. A statistical significance was set at $p = 0.05$.

To categorise the measures of performance according to their complexity, we performed two analyses. In the first one, we categorised the measures of performance found in the studies as psychomotor, perceptual, and cognitive [23,35], while in the second, we used a simple and complex categorisation: Psychomotor and perceptual measures were assigned to the simple group and cognitive measures to the complex group [23]. The categorisation was made according to the description of the tasks provided in the studies. We categorised data points ($\lambda_{\%}$) into speed or accuracy depending on the performance metric used. Table B1 in Appendix B shows the different metrics found and how they were classified. Finally, the climate was classified following the Köppen-Geiger climate categorisation [59]. We classified the data points ($\lambda_{\%}$) by the main climate groups (i.e., Tropical, Dry, Temperate, and Continental) according to the city or region where the study was performed. Due to the limited number of data points, data was not further subdivided beyond the main climate groups. When a study did not provide details of its location, we asked the authors for the information or used the location of the first author's institutional affiliation.

3. Results

3.1. Database

Thirty-five studies found in 29 peer-reviewed journal publications met the inclusion criteria for the present meta-analysis (Table 2). Some of the authors reported more than one study in their papers. Research by Chiles [60], Niemelä et al. [61], and Tham and Willem [62] each presented two studies, and Pepler reported four studies [63]. Different experiments reported in the same paper were considered to be independent since they engaged different groups of participants. Conversely, different environmental conditions with the same participant groups were considered as a single experiment. The literature review process detailing the number of studies excluded and the reason for exclusion can be seen in Fig. 2. The publication dates of the selected articles range from 1946 to 2019, which covered nearly seventy-five years of research. The journals in which studies were identified cover a wide range of disciplines, such as energy, buildings, environmental engineering, human factors, ergonomics, gerontology, psychology, and neuroscience. Seven publications met the inclusion criteria but were not included in the database, because they did not report sufficient data to perform the estimation of $\lambda_{\%}$ as described in the Methods section, for example, the performance of the participants was measured, but the results were not included in the publication [51,82–87].

The reviewed studies were performed in 11 different countries considering four out of the five Köppen-Geiger main climate groups (i.e., Tropical, Dry, Temperate, and Continental) [59] (Fig. 3). More than 80% of the performance measures were carried out in Temperate and Continental climates, and only eight measures were in Dry climates. In total, 1134 adults participated, comprising 55% males and 45% females.

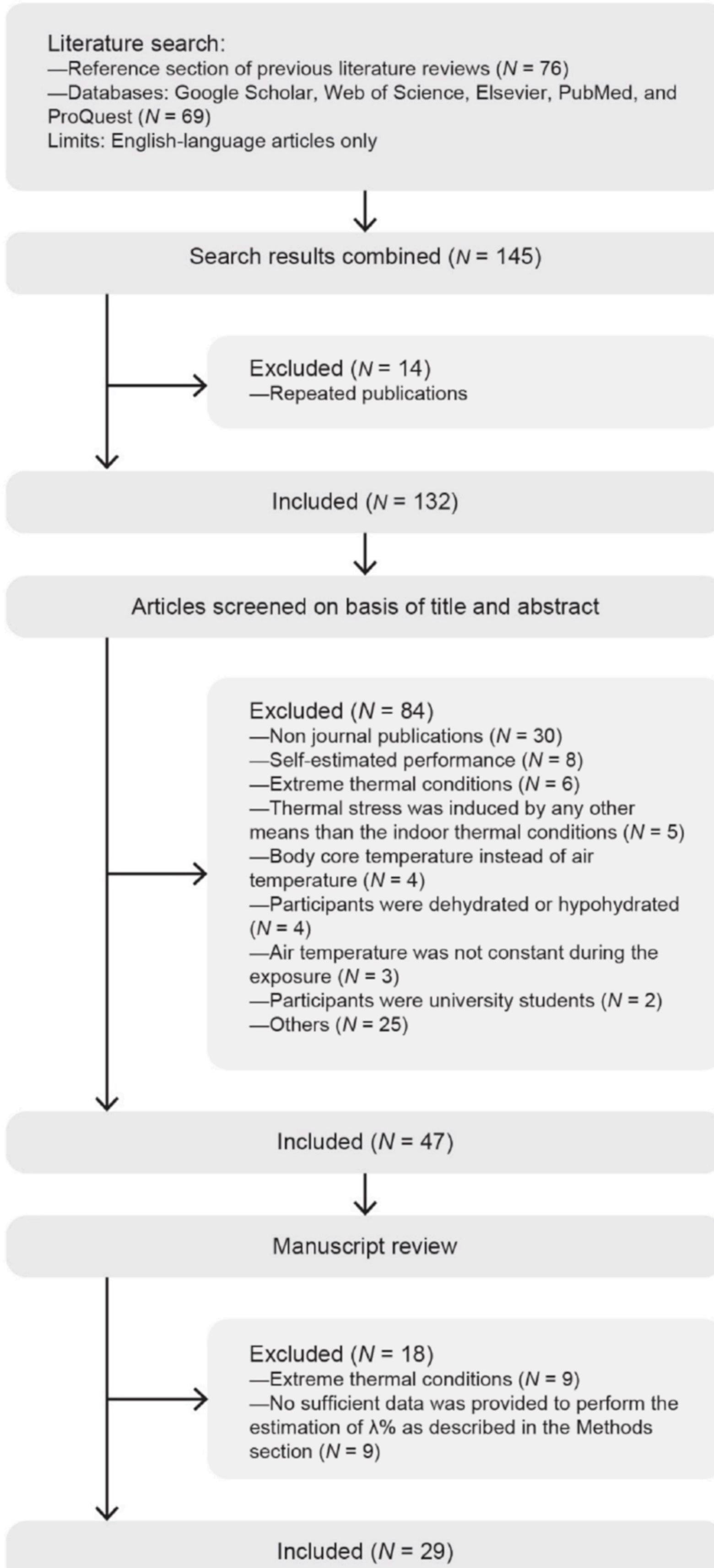

Fig. 2. Literature review process.

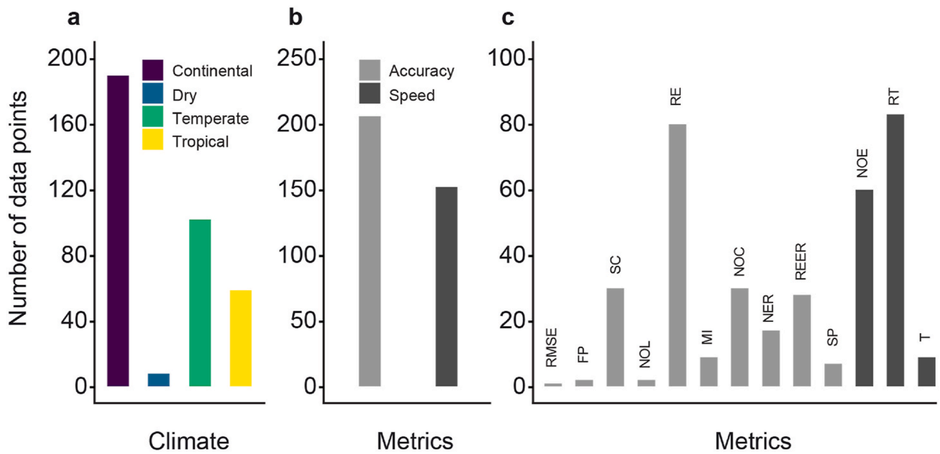

Fig. 3. Distribution of the 358 normalised data points according to climate (3a) and performance metrics (3b and 3c). Fig. 3c shows the types of metrics used: RMSE: Root Mean Square Error | FP: False positives | SC: Scores | NOL: Number of lags | RE: % of correct | MI: Missed | NOC: Number of correct | NER: Number of errors | REER: % of errors | SP: Span | NOE: Number of exercises | RT: Reaction time |T: Time.

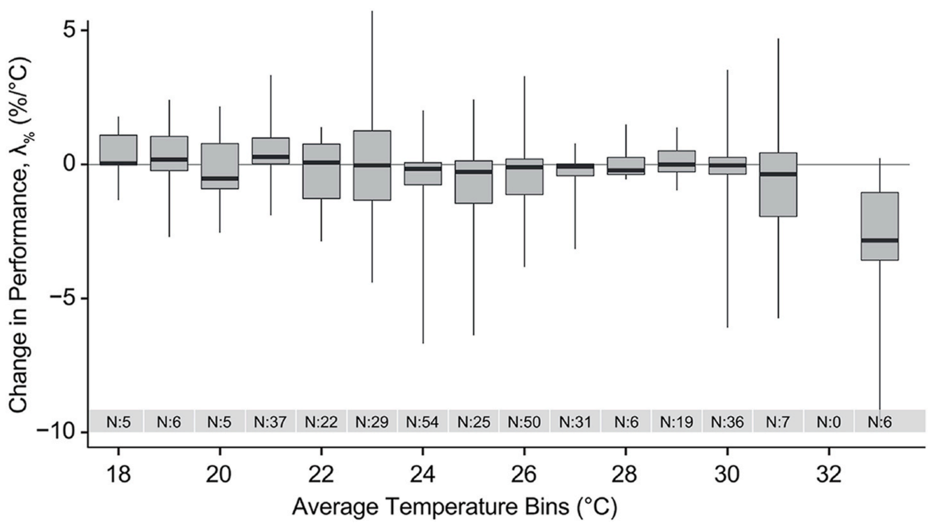

Fig. 4. Boxplot displaying the 5th, 25th, 50th, 75th, and 95th percentiles of $\lambda_{\%}$ (y-axis) for average temperature (T_{avg}) bins (x-axis) in a range that goes from 18 °C to 33 °C. N shows the number of $\lambda_{\%}$ in each T_{avg} bin.

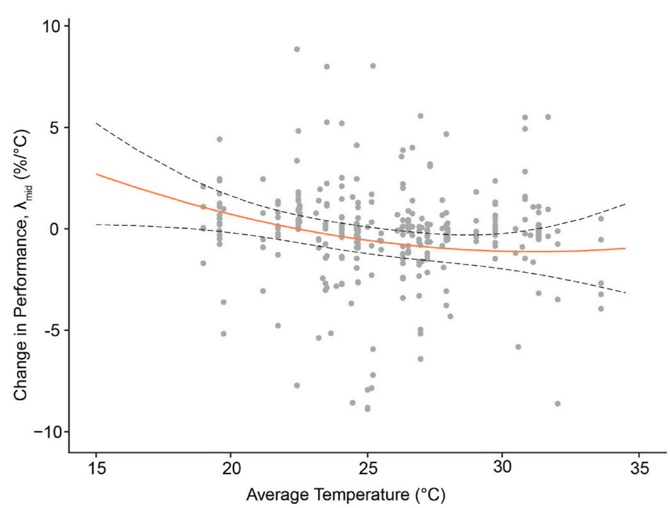

Fig. 5. Model presented in Equation (6b) (solid line) with its 95% confidence intervals (dashed lines) and the $\lambda_{\%}$ values from our database transformed into λ_{mid} (dots). The model is valid across a range of temperatures between 15 °C and 35 °C. All normalised data points (N = 358) were included in the analysis. For visualization efficiency, four data points were removed from the figure (but not from the analysis) because the λ_{mid} value was below −10%. The model's prediction power is described by $R^2 = 0.02$, $MAE = 1.7\%$, and $RMSE = 2.7\%$.

In half of the studies, participants were college or university students performing tests and tasks relevant for office work, in 40% they were office workers and soldiers, and in the remaining 10% the occupation was not reported. The median number of subjects per study was 26 persons; the biggest sample included 144 participants [69] while the smallest sample had four subjects [66]. Except for Niemelä [61] and Tham [8] —who conducted their experiments in call centres— and Bell [69] whose location was unknown, all studies were performed in controlled environments. Excluding studies that measured actual work performance [8,61], most studies introduced pre-experiment sessions to help familiarise their participants with the cognitive tasks and to minimise learning effects during the actual experiment.

Our database comprised of 571 measures of performance (i.e., the participants mean performance on a task or test under a particular thermal condition), from which we normalised 358 data points ($\lambda_{\%}$). For each task or test carried out under two different thermal conditions, we were able to calculate one $\lambda_{\%}$; however, if there were three or four conditions examined, we calculated two or three $\lambda_{\%}$ respectively. Individual performance was measured using two existing outcome metrics (e.g., talk time and telephone communications), two simulated work tasks (e.g., Morse messages and text typing), and 69 diagnostic tests including sustained attention, visual searching, working memory, inhibition, and mathematical and grammatical reasoning tasks, among others (See the open-source database for a detailed list of the tasks and tests [53]). Although the performance was measured for each participant, on many occasions the authors only reported the mean result without the standard deviation for the entire group. Participant performance was quantified by measuring the accuracy and/or speed at which they completed the assigned tasks or tests. A detailed list of these metrics can be found in Table B1 in the SM. From the 358 normalised data points ($\lambda_{\%}$), 152 (42%) correspond to speed and 206 (58%) to accuracy metrics (Fig. 3b). Of the 13 studies conducted before 2000, only Givoni and Rin [66] used both speed and accuracy metrics to assess performance, while the others used only accuracy.

We grouped all the normalised data points ($\lambda_{\%}$) into bins, for example, data points with an average temperature (T_{avg} - Equation (4)) between ≥ 17.5 °C and <18.5 °C were grouped into a bin called 18 °C, likewise average temperature between ≥ 18.5 °C and <19.5 °C were grouped into a bin called 19 °C. We continued this binning process for every increase in 1 °C until reaching the highest T_{avg} bin of 33 °C (≥32.5 °C and <33.5 °C). Fig. 4 shows boxplots for each bin. The median value was close to zero and $\lambda_{\%}$ spread both above and below this point. A wide range of change in performance ($\lambda_{\%}$) was observed between 23 °C and 26 °C, where the bins with the highest number of normalised data points are located. Similar variability in $\lambda_{\%}$ can also be observed at the 30 °C-bin and above.

3.2. Reanalysis of Seppänen et al.'s model

As described in the Methods section, we digitalised and replicated the Seppänen et al.'s model (percentage change in performance per 1 °C against temperature ($\lambda_{\%}$), in Fig. 1a) using two different datasets (the *line dataset* and the *point dataset*) and fit a quadratic polynomial model. The numerical expressions for the model fit are presented in Equation (6). There is a high correlation between both equations (r ≈ 1); therefore, we concluded that the model fit by Seppänen et al. for their unweighted data, would be similar to those presented in Equations (6a) and (6b).

Line dataset:

$$\lambda_{mid} = 13.276 - 0.935 \cdot T_{avg} + 0.015 \cdot T_{avg}^2 \tag{6a}$$

Point dataset:

$$\lambda_{mid} = 13.076 - 0.910 \cdot T_{avg} + 0.015 \cdot T_{avg}^2 \tag{6b}$$

Where T_{avg} is the average temperature in °C.

Using equation (6b), we estimated how well the model fit performed on Seppänen et al.'s data (*point dataset*). We found that the relationship between the variables was statistically significant ($p = 0.02$), but only 5% ($R^2 = 0.05$) of the variance could be explained by the average temperature (Fig. 1a). The *MAE* showed that on average there is a difference of 1.9% between the fit and the values for the observed λ_{mid}. The *RMSE* was 3.1% and since *MAE* and *RMSE* are expressed in the same units as the response variable, 1.9% and 3.1% are high errors when considering that the highest λ_{mid} predicted by the model is 2.8% at 15 °C.

We also tested how well our normalised data ($N = 358$) fit to Seppänen et al.'s model. Fig. 5 shows our normalised data points after transforming our values of $\lambda_{\%}$ into λ_{mid}. In Equation (6b), the model along with the 95% confidence intervals are presented. The R^2, *MAE*, and *RMSE* between the observed (our values) and predicted λ_{mid} (i.e., by Seppänen et al.'s model) were 0.02, 1.7% and 2.7%, respectively, which

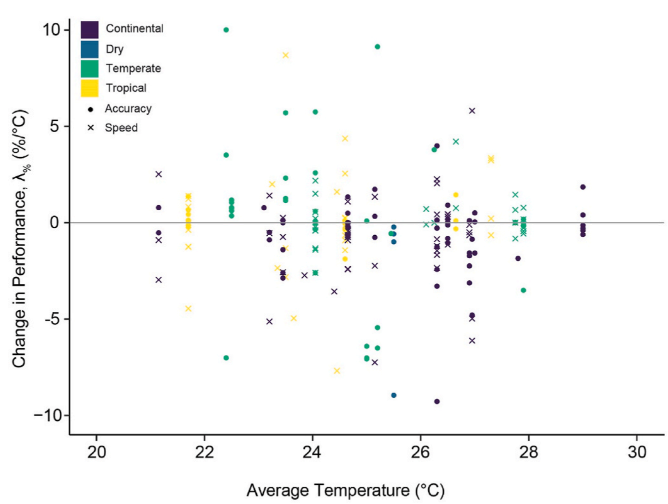

Fig. 6. The percentage change in performance per 1 °C increment in temperature ($\lambda_{\%}$) plotted against the average temperature (T_{avg}). Positive values indicate improved performance with increased temperature, negative values indicate that performance is reduced with increased temperature, and the values along the $\lambda_{\%}$ = 0 line mean no change in performance. There are 215 normalised data points (N_1) distributed in a range of average temperatures (T_{avg}) that goes from 20 °C to 30 °C.

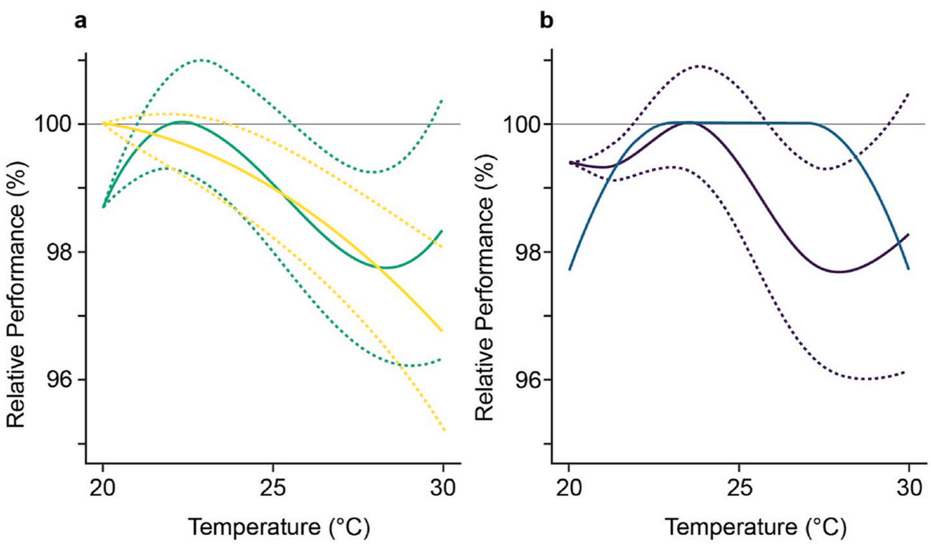

Fig. 7. Relative performance in relation to a reference temperature of 20 °C. In Fig. 7a, yellow and green lines correspond to the linear (p = 0.6, R^2 = 0.05, *MAE* = 1.6%, and *RMSE* = 2.5%) and quadratic (p = 0.29, R^2 = 0.07, *MAE* = 1.6%, and *RMSE* = 2.5%)), models, while in Fig. 7b, blue and purple lines correspond to the *LOESS* (p = NA, R^2 = 0.06, *MAE* = 1.6%, and *RMSE* = 2.5%) and Maximal Adaptability (*MAM*) models (R^2 = 0.0, *MAE* = 1.6%, and *RMSE* = 2.6%). For the linear, quadratic, and *LOESS* models, solid lines represent the median and dashed lines the 2.5th and the 97.5th percentiles curves estimated as specified in the Methods section and Appendix F. The p-value, R^2, *MAE*, and *RMSE* correspond to the average values of the performance metrics of 1000 models. For the *MAM*, a single model was fit and the R^2, *MAE*, and *RMSE* correspond to the performance metrics (See Appendix I). **Note 1:** The performance metrics provided with the models are associated with the relationships between the percentage change in performance and the average temperature presented in Appendix E, but were included here because Fig. 7a and b are graphical transformations of those relationships. **Note 2:** Due to the low accuracy of the models, no conclusions should be drawn from the shape of the models. Despite having different shapes, the objective of presenting them here is to show how they all have a similar accuracy and are comparable to the zero-order model. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

reveals a low predictive capacity and a small proportion of the variance in the performance that can be explained by the average air temperature. The results show that when the original [20] or new data collected are used, Seppänen et al.'s model has a very low predictive power when determining the percentage change in work performance (λ_{mid}). Furthermore, we also fit a zero-order model and obtained a *MAE* of 2.0% and *RMSE* of 3.2% which is comparable to the Seppänen et al.'s model.

3.3. Prediction models for the percentage change in work performance

Fig. 6 shows the percentage change in performance per 1 °C increment in temperature ($\lambda_{\%}$) plotted against the average temperature (T_{avg} – Equation (4)). Given that 90% of office buildings are within an indoor temperature range between 20 °C and 30 °C [52], we decided to only include measures of performance from studies when temperature conditions were within this range. Consequently, we removed 48 data points when $T_L < 20$ °C and 95 data points when $T_H > 30$ °C. The graph includes 215 (N_1) normalised measures of performance classified by the metric (i.e., speed or accuracy) and climate zone.

Using average temperature (T_{avg}) in our database as the predictor, we applied different models to predict the change in performance per 1 °C. Table 3 summarises the models that we introduced for the prediction analysis and the corresponding statistical metrics (i.e., R^2, *MAE*, and *RMSE*) for each model. These metrics reflect how well the model will perform on the new data and generally, we found a low prediction power across for all models, and for each metric, there was a small range in the level of prediction (R^2 = 0.0–0.07; *MAE* = 1.6%–1.8%; and *RMSE* = 2.4%–2.8%). Despite the proportion of variance explained by two of the machine learning algorithms including k-Nearest Neighbours and Random Forest (R^2 = 0.07), these models present characteristics of overfitting (Appendix D in the SM). Since these models are likely to describe the random error in the data rather than the relationship between $\lambda_{\%}$ and T_{avg}, we discarded their use in the subsequent analyses.

In 1986, Hancock and Warm [88] proposed the Maximal Adaptability Model (*MAM*) that relates thermal stress to performance. Unlike the inverted-U model (the base of the Seppänen et al.'s relationship [20]), where maximum performance is achieved at a single temperature (or thermal sensation), the extended-U model proposes that human performance remains relatively stable over a wide range of temperatures but deteriorates rapidly outside this range. Although its use in the field of indoor environmental science has been limited [34], the *MAM* is a widely known model in other disciplines. Similar to what was applied to the model of Seppänen et al., we assessed its accuracy by including a model in Table 3 that follows the extended-U model and validated it with our normalised data (N_1 = 215). The detailed procedure can be found in Appendix I in SM. The results show that the prediction performance of the *MAM* does not differ much compared to Seppänen et al.'s model or other investigated in this study: R^2 = 0.0, *MAE* = 1.6%, and *RMSE* = 2.6%.

All remaining models had a poor prediction power similar to the zero-order model, which assumes temperature does not affect $\lambda_{\%}$ and its value is constant and equal to the mean $(\overline{\lambda_{\%}})$. Our results suggest that, regardless of the prediction model choice, using only the average air temperature (T_{avg}) had negligible significance in predicting the percentage change in performance.

Despite the low accuracy for the metrics presented in Table 3, we have chosen four models (i.e., linear, quadratic, *LOESS*, and *MAM*) to be used in subsequent analyses. As described in the Methods section, for the confidence range analysis we randomly created 1000 samples for three of the selected models (i.e., linear, quadratic, and *LOESS*. Details of *MAM* model developing can be seen in Appendix I) with different training and testing datasets. Therefore, 3000 sets of training and testing data were generated and used to fit an equal number of models (1000 for linear, 1000 for quadratic polynomial, and 1000 for *LOESS*), which in turn, were used to predict the values of $\lambda_{\%}$ in a range of temperatures between 20 °C and 30 °C with a step-size of 1 °C. The results were used to calculate the median, 2.5th and 97.5th percentiles at every temperature level. A graphical representation of these relationships and the performance metrics of the model can be seen in Appendix E in the SM. We also transformed these relationships into a more useable graph quantifying the relationship between the relative performance (*RP*) and temperature. An example of the transformation procedure is presented in Appendix F.

Fig. 7 shows the transformed relationship between *RP* and temperature for the linear, quadratic, *LOESS*, and *MAM* models. These models produced a wide variety of relationships and this exemplifies how model selection has a direct impact on the final relationship between temperature and performance, as well as the temperature at which maximum performance is reached. In the examples given in Fig. 7, maximum performance is achieved across a wide range of temperatures with differences of up to 4 °C and this also occurs in Seppänen et al.'s model. Although the R^2 of the linear model is <1% lower than the quadratic fit, if Seppänen et al. had chosen the linear approach instead of the quadratic model, the optimal temperature at which maximum performance was observed would be 23.1 °C instead of the 21.8 °C.

3.4. Multiple regression analysis

We created dummy variables for each of the categorical variables (See Appendix G in the SM). Continental climate, cognitive complexity, and accuracy were used as the reference factors because they were the categories that presented the highest number of observations. We used a dataset with 211 normalised data points (N_2), excluding the four data points observed in the dry climate condition. The multiple regression analysis showed that the coefficients associated with temperature or with any of the dummy variables included were not statistically significant ($p > 0.05$). Thus, we do not show here any models fit to a subset of data points based on their category, for example, models fit to speed or accuracy measures. However, model fitting by category can be explored via the online application that we have developed (http://cbe-berkeley.shinyapps.io/temp-performance).

4. Discussion

We collected the available evidence on the effects of temperature on office work performance and examined the psychomotor, perceptual, and cognitive skills relevant to office work. We normalised 571 performance measures obtained in 35 studies by calculating the percentage change in performance per 1 °C increase in temperature ($\lambda_{\%}$). We created an open-source database with 358 normalised data points ($\lambda_{\%}$). We reanalysed the relationship between performance and temperature developed by Seppänen et al. [20] and by using a subset of our database, performance measures were examined using a range of temperatures between 20 °C and 30 °C.

We explored the possibility to develop a new relationship via different prediction models including regression models, machine learning algorithms, and models based on the Maximal Adaptability framework. The results show that both the model proposed by Seppänen et al. [20] and the models developed by us have a prediction power so low that they could not be used. When applying the original data from Seppänen et al. [20] and the data from our database to Seppänen et al.'s model, we found that temperature could explain only 5% ($MAE = 1.9\%$, $RMSE = 3.1\%$) and 2% ($MAE = 1.7\%$ and $RMSE = 2.7\%$) of the variance in $\lambda_{\%}$, respectively. In the case of our developed models (see Table 3), the level of prediction was always not statistically significant and showed poor performance accuracy. The maximum R^2 was 7%, and the lowest *MAE* and *RMSE* were 1.6% and 2.5%, respectively. The performance metrics from all these models are comparable to the zero-order model, which assumes there is no relationship between work performance and temperature.

Given that only a very small proportion of the variance in the response can be explained by the temperature and that the models are no better than the zero-order model, we explored other factors that may explain the poor prediction accuracy, including (i) task complexity, (ii) speed and accuracy, and (iii) climate. We performed multiple linear regression analysis, but none of the factors could explain the lack of relationship ($p > 0.05$). Even when exposure time has been considered an important factor for evaluating performance, we decided not to include it in our multiple regression analysis. The first reason was due to the difficulty of defining the limits between long and short-term exposures. For example, in their literature reviews, Yeganeh et al. [23] used the mean task time of the selected studies as the threshold between short- and long-term exposures; conversely, Ramsey [38] designated short-term exposures as those less than 30 min, while those between 4 and 8 h were classified as long-term. However, these categorisations correspond to the authors' judgements without justification in the scientific literature. The second reason is that in our case, except for the studies carried out by Tham [8] and Tham and Willem [62] in call centres and the five studies that did not report the exposure time, all the other 27 reported an exposure time within a very short range (i.e., 25–270 min, mean = 140 ± 60 min). Nevertheless, Ramsey [38] found no effects of the duration of exposure on performance within a exposure range of up to 400 min.

Our findings in this study are different from previously published works [4,9,12,17–23], which had found relationships between temperature and performance within a temperature range similar to ours (20 °C–30 °C). Relationships reported in former studies include: Seppänen et al. [20] and Lan et al. [12] that showed a decrease of 8–9% in performance when temperature increased from 22 °C to 30 °C; Pilcher et al. [18] that identified small effects on performance at wet bulb globe temperatures (WBGT) between 21 °C and 27 °C. When assuming that relative humidity was 60% and dry-bulb and globe temperatures are equal, these temperatures were equivalent to 25 °C and 31 °C; Geng et al. [22] that showed a 5% increase in performance at 26 °C. However, performance at 22 °C and 29 °C were almost identical; and Yeganeh et al. [23] reported a 0.4% decrease in performance when temperature increased from ~22 °C to ~26 °C. The variety of analytical methods in these studies, the multiple confounders, the lack of prediction accuracy metrics (as discussed below), and our normalised data points representing the average performance instead of individual performance, are plausible reasons for these differences.

Because other studies —except from Geng et al. [22] and Cui et al. [9]— did not provide measures of model accuracy, we could not compare our models directly with the previously published models. Geng et al. [22] built their relationship from a database containing 21 participants and reported an R^2 of 0.9 when the temperature was used as the predictor variable. Similarly, Cui et al. [9] developed their relationship using a database of 36 participants and reported an R^2 of 0.99. However, these authors did not fit their models based on the performance of individuals but took an average value of performance at each of the investigated temperatures (seven in the former and five in the latter study). This averaging procedure artificially minimises the variance in the model, thus increases the R^2 value.

We further evaluated model prediction performance in Table 3 by considering the wider range of temperatures that were available in the database (i.e., 18 °C–34 °C, $N = 358$ data points). The results showed that even though the models performed slightly better (see a summary table in Appendix H in SM), prediction power was still low ($R^2 = 0.04$–0.10; $MAE = 1.5\%$–1.7%; and $RMSE = 2.4\%$–2.7%) and was similar to the zero-order model ($MAE = 1.6\%$ and $RMSE = 2.5\%$). We followed the same procedure with the MAM and fit several piecewise models that followed the framework of the extended-U theory. Detailed procedures can be found in Appendix I in SM. Results showed that the model with the best performance could only explain 1% of the variance in $\lambda_{\%}$, and the *MAE* and *RMSE* were 1.4% and 2.5%, which is similar to the Seppänen et al.'s model and our models in Appendix H. This implies that the inverted U-models or extended U-models (*MAM*) cannot predict the effect of temperature as an input variable on work performance when using our normalised data points.

Since all explored models in this study have a prediction performance similar to the zero-order model ($R^2 = 0$ and $\lambda_{\%} = \overline{\lambda_{\%}}$), we were unable to find a relationship between temperature and office work performance when considering the range of temperatures typically found in office buildings (20 °C–30 °C), or across a wider temperature range within our collected database (18 °C–34 °C). Consequently, we do not recommend applying in practice the models examined in the present study that only rely on temperature as a predictor of work performance. Even though we could not find the relationship, this does not preclude that such relationship(s) may exist and this should be investigated in future studies and analyses.

5. Limitations and future work

The present study put together a heterogeneous group of individual studies. There are consequently limitations on how the data can be used and analysed including the method chosen to normalise the performance outcomes, the use of temperature as a proxy of the indoor thermal environment, and the classification of task complexity. The main limitations are summarised below.

There are multiple methods to assess individual job performance [89]. Most of them were developed in the fields of industrial and organisational psychology to make a selection decision, for example, for hiring, promotion, or retention of workers. Each method has its strengths and weaknesses [89]. While task performance of manual workers can be measured based on the output or throughput, there are no universally accepted methods to measure knowledge workers' performance [90] and the assessment is more complex [91]. Office work implies a complex set of overlapping tasks, it is not homogeneous, varies with time, and requires the interpretation of information from many fields of knowledge. In the papers selected for the present analyses, a wide variety of methods were used to assess the abilities and skills needed to perform cognitive and manual tasks typical in office work. The tests had various complexities and measured a wide number of skills and mental processes using very different absolute scales and units. Moreover, each cognitive task may be affected differently by temperature which also depends on arousal. For example, Liu et al.'s study [79] showed that rule-based logical thinking (e.g., addition and Baddeley test) and tasks requiring rapid response to expected signals (e.g., tests measuring simple and multiple reaction time (RT), and redirection test) were better performed at the cool end of the acceptable temperature range when arousal is raised. However, the tasks requiring open-ended and creative thinking and vigilance for unexpected signals (e.g., creativity tests examining the alternate uses of different items) were performed better at, or slightly above, the centre of the acceptable temperature range when arousal decreased as shown by Wyon et al. [92] and Tham and Willem [75]. Taking all these considerations and differences into account, the data used in the present analysis comprised of tasks and tests where performance at different temperatures might not be directly comparable. Rather than developing a universal relationship between temperature and work performance, these differences imply that future studies should explore the feasibility of developing individual relationships for clusters of tasks that require similar cognitive abilities or skills (e.g., Ramsey and Morrissey's approach [15]), or those that are similarly affected by the thermal environment.

In our meta-analysis, the temperature was the only variable used to characterise the indoor thermal environment. Since it was the most frequently reported proxy in the selected studies, it was used in our work. However, we have demonstrated in this study that temperature alone may not be sufficient in characterising the overall thermal environment perceived by the occupant. Regardless of air temperature, combinations of other typical thermal parameters, i.e., mean radiant temperature, relative humidity, airspeed, clothing insulation and activity level, may result in similar or very different thermal sensations perceived by the occupants. Studies by Lan et al. [12,93] show that both thermal discomfort and temperature itself may affect cognitive performance. Wyon et al. [92] and Lan et al. [93] found that when the temperature was $\geq$18 °C and <27 °C [93,94] and thermal comfort is achieved, there were no negative effects on participant performance [94], while negative effects on performance were found even though thermal comfort was still achieved when temperature increased $\geq$27 °C [93]. Negative effects may also occur when people experience thermal discomfort regardless of the temperature [12].

A further caveat of the analyses performed is that the diversity of individual preferences and responses to the thermal environment may not be explained by any relationship at the group level. The group-level relationships inform building operators and users: what is the likely temperature setpoint based on work performance across all tasks and individual building occupants. They apply to large groups of people doing a range of tasks. If personal optimal performance exists, it may require a "last mile" adjustment at the individual level.

As we did not have access to the raw data, we only used the information reported in the studies. Since many studies did not report the standard deviation, the effect size could not be determined. By using Seppänen et al.'s [20] approach to normalising the data, each of our data points represents the average performance of all the participants across our selected studies. There might be variability in the group's performance that our approach prevents us from observing and could not be accounted for. To ensure research can be reproduced, future studies should provide access to the raw data (e.g., Ref. [87]).

The change in performance per 1 °C increase in temperature ($\lambda_{\%}$) used for normalising the data was regressed against the average temperature, which may have introduced errors associated with the assumption of linearity over the range of temperature differences. The normalisation approach used by Seppänen et al. [20] accepts that performance changes linearly with temperature within the range of temperatures examined, regardless of the magnitude of the temperature difference and also the level of temperatures. This conflicts with the theories concerning the effects of heat stress on cognitive performance, which state that if the stress level increases, the performance will increase until a certain point (arousal theory) or a plateau (maximal adaptability model), after which it will begin to decrease. Therefore, the linearity assumption may not always be justified. It is reasonable to assume validity only applies for the small temperature differences and becomes less reliable as the temperature difference increases. The linearity assumption has been applied to normalise all performance measures, which may not be valid either.

The categorisation of task complexity depends on the type of the task rather than the skill level required to perform the task. For example, mathematical tasks are generally classified as complex, irrespective of the cognitive abilities of the participant or the specific difficulty of the exercises (e.g., adding one-digit or multiplying three-digit numbers). For the categorisation applied in the present analyses, we assumed that cognitive tasks are more complex than perceptual tasks, which in turn are more complex than psychomotor tasks. Similar assumptions were also made by others [23,35]. Brain imaging techniques like functional magnetic resonance imaging (fMRI) or electroencephalography (EEG) have begun to show that mental processes activate different areas of the brain with different levels of intensity, hence the categorisation we used may not be accurate. In the future, the primary brain region(s) activated by a certain task can instead be used for task categorisation.

The limitations described above, seem to be similar to the ones presented in previous reviews and meta-analyses. However, the inclusion of new studies in this review and the use of heterogeneous analytical methods brought about more challenges exacerbated by these limitations. We see four future research pathways that could lead to a better understanding of the aggregated average effects within the indoor environment on office work performance, especially in the range of thermal conditions close to the thermal neutrality:

(1) Using other methods to aggregate data to determine if there is a relationship between other proxies of the indoor thermal environment, for example, thermal sensation, and office work performance. As we have presented in Table 1, some studies developed a relationship between thermal sensation and work performance, but none of them analysed the data using a large number of studies coming from different research groups. Although the thermal effect will be better characterised, the main constraint of this approach is that most of the potential confounding factors mentioned above remain. Since the quantification of people's thermal sensation had certain limitations [95] and has been used less than temperature when work performance

was measured, the number of research studies and available data points could be lower —further limiting this approach.

(2) The use of raw data from different studies. Performing a meta-analysis with raw data can help control the variance and clarify where the variability in performance is originating from. It can also unlock other paths to regroup or normalise the data using, for example, the effect size. However, there still will be factors that cannot be controlled such as the heterogeneity of methods used to assess performance and the research protocol.

(3) Development of a protocol describing how experiments linking indoor thermal conditions to office work performance should be performed and what should be reported in the publications, for example, the skills that should be measured, the assessment methods and metrics that should be used, open-source access to raw data, reporting the statistical metrics of the measured data, etc. To this end, Yeganeh et al. [23] mentioned that "… [studies] could become more conclusive through improved study methods and by providing more detailed study statistics for meta-analysis, generalization, and model building". The establishment of such a protocol makes reproducible research possible, which further consolidates the findings reported in future studies.

(4) Conducting a large-scale experiment across different locations. Such experiment should have control over the major confounders, especially implementing standardised protocols across all conditions and experiments and making the raw data accessible (as suggested in (3)). The main limitations of this approach are resources and that the work should be carried out as a collaborative effort across multiple research groups. Also, many of the limitations and confounding factors mentioned above will still be valid.

6. Conclusions

Through a meta-analysis, we put together an open-source database examining 35 studies on the effects of temperature on office work performance, and psychomotor, perceptual and cognitive abilities essential for office work. The database comprises 358 normalised data points, which even when considering the limitations discussed above, makes it the most comprehensive and complete database that has been developed to date.

We showed that the model developed by Seppänen et al. [20] should not be used to predict the effect of temperature on work performance because of its low prediction power, which was validated using original data from Seppänen et al.: $R^2 = 0.05$, $MAE = 1.9\%$, $RMSE = 3.1\%$, and the data collected in the present study: $R^2 = 0.02$, $MAE = 1.7\%$, $RMSE = 2.7\%$. This means that only a very small proportion of variance corresponding to work performance is explained by the temperature (i.e., the predictor variable). We also reviewed other models describing the relationship between thermal effects and performance (Table 1). Common observations we see across these studies are a lack of statistical information reported and the raw data is not made open-source. We argue that future studies and modelling exercises should provide this information as per our recommendations suggested in the previous section (experimental protocol).

Using the collected data, we examined different regression models, models based on the Maximal Adaptability framework, and also machine learning algorithms, but we still could not find a relationship between temperature and office work performance both across the entire temperature range available in the database (18 °C–34 °C) or for the typical temperature range in office buildings (20 °C–30 °C). Using multiple regression analyses, we found that factors such as task complexity, metrics of performance such as speed and accuracy, and data coming from different climatic regions did not show a significant impact on the relationship between temperature and office work performance.

These results do not necessarily imply that there are no effects between thermal environments in buildings and office work performance. They indicate that given limitations such as the normalisation procedure and analytical methods, multiple uncontrolled confounders, and temperature as the sole predictor variable, we could not find a relationship between temperature and work performance. As long as the raw data is open access and the sources of variability are revealed, some of the limitations could be addressed in a future meta-analysis. However, even if we could arrive at a relationship with sufficient accuracy to be used in practice, it would only be a relationship that applies to the group level (i.e., across all tasks and individual building occupants), meaning that we still cannot explain the diversity of individual preferences and responses to the thermal environment.

In conclusion, we do not recommend applying in practice the models examined in the present study that only rely on temperature as a predictor of work performance —including the one developed by Seppänen et al. Coordinated research predicated on a shared protocol enabling integrated analysis in the modelling of the relationships between the indoor thermal environment and office work performance is needed before using these relationships in practice.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was funded by the Republic of Singapore's National Research Foundation through a grant to the Berkeley Education Alliance for Research in Singapore (BEARS) for the Singapore-Berkeley Building Efficiency and Sustainability in the Tropics (SinBerBEST) Program. BEARS has been established by the University of California, Berkeley as a centre for intellectual excellence in research and education in Singapore. Dr Michael Kent is specially thanked for his comments on this manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.buildenv.2021.108037.

References

[1] K. Parsons, Human Thermal Environments: the Effects of Hot, Moderate, and Cold Environments on Human Health, Comfort, and Performance, Third, CRC Press, London, UK, 2014.

[2] J.E. Woods, Cost avoidance and productivity in owning and operating buildings, Occup. Med. 4 (1989) 753–770.

[3] O. Seppänen, Estimated cost of indoor climate in Finnish buildings, in: Proc. Eighth Int. Conf. Indoor Air Qual. Clim., 1999, pp. 13–18. Edinburgh, Scotland, https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Estimated+cost+of+indoor+climate+in+Finnish+buildings&btnG=. (Accessed 5 November 2019).

[4] P. Roelofsen, The impact of office environments on employee performance: the design of the workplace as a strategy for productivity enhancement, J. Facil. Manag. 1 (2002) 247–264.

[5] P. Wargocki, D.P. Wyon (Eds.), Indoor Climate and Productivity in Offices: How to Integrate Productivity in Life-Cycle Cost Analysis of Building Services, Rehva, Brussels, 2006.

[6] W.J. Fisk, A.H. Rosenfeld, Estimates of improved productivity and health from better indoor environments, Indoor Air 7 (1997) 158–172, https://doi.org/10.1111/j.1600-0668.1997.t01-1-00002.x.

[7] W.J. Fisk, D. Black, G. Brunner, Benefits and costs of improved IEQ in U.S. offices: benefits and costs of improved IEQ in U.S. offices, Indoor Air 21 (2011) 357–367, https://doi.org/10.1111/j.1600-0668.2011.00719.x.

[8] K.W. Tham, Effects of temperature and outdoor air supply rate on the performance of call center operators in the tropics, Indoor Air 14 (2004) 119–125, https://doi.org/10.1111/j.1600-0668.2004.00280.x.

[9] W. Cui, G. Cao, J.H. Park, Q. Ouyang, Y. Zhu, Influence of indoor air temperature on human thermal comfort, motivation and performance, Build. Environ. 68 (2013) 114–122, https://doi.org/10.1016/j.buildenv.2013.06.012.

[10] S. Tanabe, N. Nishihara, Productivity and fatigue, Indoor Air 14 (2004) 126–133, https://doi.org/10.1111/j.1600-0668.2004.00281.x.

[11] S. Lamb, K.C.S. Kwok, A longitudinal investigation of work environment stressors on the performance and wellbeing of office workers, Appl. Ergon. 52 (2016) 104–111, https://doi.org/10.1016/j.apergo.2015.07.010.

[12] L. Lan, P. Wargocki, Z. Lian, Quantitative measurement of productivity loss due to thermal discomfort, Energy Build. 43 (2011) 1057–1062, https://doi.org/10.1016/j.enbuild.2010.09.001.

[13] T. Nayak, T. Zhang, Z. Mao, X. Xu, L. Zhang, D. Pack, B. Dong, Y. Huang, Prediction of human performance using electroencephalography under different indoor room temperatures, Brain Sci. 8 (2018) 74, https://doi.org/10.3390/brainsci8040074.

[14] X. Wang, D. Li, C.C. Menassa, V.R. Kamat, Investigating the effect of indoor thermal environment on occupants' mental workload and task performance using electroencephalogram, Build. Environ. 158 (2019) 120–132, https://doi.org/10.1016/j.buildenv.2019.05.012.

[15] J.D. Ramsey, S.J. Morrissey, Isodecrement curves for task performance in hot environments, Appl. Ergon. 9 (1978) 66–72, https://doi.org/10.1016/0003-6870(78)90150-3.

[16] D.P. Wyon, The effects of indoor climate on productivity and performance: a review (in Swedish), WS Energi 3 (1986) 59–65.

[17] L. Berglund, R. Gonzales, A. Gagge, Predicted human performance decrement from thermal discomfort and ET*, in: Proc. Fifth Int. Conf. Indoor Air Qual. Clim. Tor. Can., 1990, pp. 215–220.

[18] J.J. Pilcher, E. Nadler, C. Busch, Effects of hot and cold temperature exposure on performance: a meta-analytic review, Ergonomics 45 (2002) 682–698, https://doi.org/10.1080/00140130210158419.

[19] R. Kosonen, F. Tan, Assessment of productivity loss in air-conditioned buildings using PMV index, Energy Build. 36 (2004) 987–993, https://doi.org/10.1016/j.enbuild.2004.06.021.

[20] O. Seppänen, W.J. Fisk, Q.H. Lei, Effect of temperature on task performance in offfice environment, in: Proc. 5th Int. Conf. Cold Clim. Vent. Air Cond., 2006, p. 11. Moscow, Russia.

[21] K.L. Jensen, J. Toftum, P. Friis-Hansen, A Bayesian Network approach to the evaluation of building design and its consequences for employee performance and operational costs, Build. Environ. 44 (2009) 456–462, https://doi.org/10.1016/j.buildenv.2008.04.008.

[22] Y. Geng, W. Ji, B. Lin, Y. Zhu, The impact of thermal environment on occupant IEQ perception and productivity, Build. Environ. 121 (2017) 158–167, https://doi.org/10.1016/j.buildenv.2017.05.022.

[23] A.J. Yeganeh, G. Reichard, A.P. McCoy, T. Bulbul, F. Jazizadeh, Correlation of ambient air temperature and cognitive performance: a systematic review and meta-analysis, Build. Environ. 143 (2018) 701–716, https://doi.org/10.1016/j.buildenv.2018.07.002.

[24] O. Seppänen, W.J. Fisk, Q.H. Lei, Room Temperature and Productivity in Office Work, Lawrence Berkeley National Laboratory, California, USA, n.d.

[25] O. Seppanen, W. Fisk, Some quantitative relations between indoor environmental quality and work performance or health, HVAC R Res. 12 (2006) 957–973, https://doi.org/10.1080/10789669.2006.10391446.

[26] T. Deryugina, S. Hsiang, Does the Environment Still Matter? Daily Temperature and Income in the United States, National Bureau of Economic Research, Cambridge, MA, 2014, https://doi.org/10.3386/w20750.

[27] M. Dell, B.F. Jones, B.A. Olken, What do we learn from the weather? The new climate-economy literature, J. Econ. Lit. 52 (2014) 740–798, https://doi.org/10.1257/jel.52.3.740.

[28] G. Heal, J. Park, Feeling the Heat: Temperature, Physiology & the Wealth of Nations, National Bureau of Economic Research, Cambridge, MA, 2013, https://doi.org/10.3386/w19725.

[29] J. Park, J. Goodman, M. Hurwitz, J. Smith, Heat and learning, Am. Econ. J. Econ. Pol. 12 (2020) 306–339, https://doi.org/10.1257/pol.20180612.

[30] R. Valancius, A. Jurelionis, V. Dorosevas, Method for cost-benefit analysis of improved indoor climate conditions and reduced energy consumption in office buildings, Energies 6 (2013) 4591–4606, https://doi.org/10.3390/en6094591.

[31] P.M. Bluyssen, What do we need to be able to (re)design healthy and comfortable indoor environments? Intell. Build. Int. 6 (2014) 69–92, https://doi.org/10.1080/17508975.2013.866068.

[32] F. Mofidi, H. Akbari, Integrated optimization of energy costs and occupants' productivity in commercial buildings, Energy Build. 129 (2016) 247–260, https://doi.org/10.1016/j.enbuild.2016.07.059.

[33] M.S. Owen (Ed.), Ashrae Handbook: Fundamentals: SI Edition, American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., Atlanta, GA, 2017.

[34] F. Zhang, R. de Dear, P. Hancock, Effects of moderate thermal environments on cognitive performance: a multidisciplinary review, Appl. Energy 236 (2019) 760–777, https://doi.org/10.1016/j.apenergy.2018.12.005.

[35] P.A. Hancock, J.M. Ross, J.L. Szalma, A meta-analysis of performance response under thermal stressors, Hum. Factors J. Hum. Factors Ergon. Soc. 49 (2007) 851–877, https://doi.org/10.1518/001872007X230226.

[36] P.A. Hancock, I. Vasmatzidis, Effects of heat stress on cognitive performance: the current state of knowledge, Int. J. Hyperther. 19 (2003) 355–372, https://doi.org/10.1080/0265673021000054630.

[37] F. Zhang, R. de Dear, University students' cognitive performance under temperature cycles induced by direct load control events, Indoor Air 27 (2017) 78–93, https://doi.org/10.1111/ina.12296.

[38] J.D. Ramsey, Task performance in heat: a review, Ergonomics 38 (1995) 154–165, https://doi.org/10.1080/00140139508925092.

[39] L. Taylor, S.L. Watkins, H. Marshall, B.J. Dascombe, J. Foster, The impact of different environmental conditions on cognitive function: a focused review, Front. Physiol. 6 (2016), https://doi.org/10.3389/fphys.2015.00372.

[40] S. Qian, G. Sun, Q. Jiang, K. Liu, B. Li, M. Li, X. Yang, Z. Yang, L. Zhao, Altered topological patterns of large-scale brain functional networks during passive hyperthermia, Brain Cognit. 83 (2013) 121–131, https://doi.org/10.1016/j.bandc.2013.07.013.

[41] M. Chignell, T. Tong, S. Mizobuchi, W. Walmsley, Combining speed and accuracy into a global measure of performance, Proc. Hum. Factors Ergon. Soc. Annu. Meet. (2014) 1442–1446, https://doi.org/10.1177/1541931214581301.

[42] R. Bogacz, E. Wagenmakers, B.U. Forstmann, S. Nieuwenhuis, The neural basis of the speed–accuracy tradeoff, Trends Neurosci. 33 (2010) 10–16, https://doi.org/10.1016/j.tins.2009.09.002.

[43] R.P. Heitz, The speed-accuracy tradeoff: history, physiology, methodology, and behavior, Front. Neurosci. 8 (2014), https://doi.org/10.3389/fnins.2014.00150.

[44] W.A. Wickelgren, Speed-accuracy tradeoff and information processing dynamics, Acta Psychol. (Amst.). 41 (1977) 67–85, https://doi.org/10.1016/0001-6918(77)90012-9.

[45] N.A.S. Taylor, Human heat adaptation, in: R. Terjung (Ed.), Compr. Physiol., John Wiley & Sons, Inc., Hoboken, NJ, USA, 2014, pp. 325–365, https://doi.org/10.1002/cphy.c130022.

[46] P. Wargocki, J.A. Porras-Salazar, S. Contreras-Espinoza, The relationship between classroom temperature and children's performance in school, Build. Environ. 157 (2019) 197–204, https://doi.org/10.1016/j.buildenv.2019.04.046.

[47] S. Tanabe, M. Haneda, N. Nishihara, Workplace productivity and individual thermal satisfaction, Build. Environ. 91 (2015) 42–50, https://doi.org/10.1016/j.buildenv.2015.02.032.

[48] M.J. Mendell, G.A. Heath, Do indoor pollutants and thermal conditions in schools influence student performance? A critical review of the literature, Indoor Air 15 (2005) 27–52, https://doi.org/10.1111/j.1600-0668.2004.00320.x.

[49] M. Dawe, P. Raftery, J. Woolley, S. Schiavon, F. Bauman, Comparison of mean radiant and air temperatures in mechanically-conditioned commercial buildings from over 200,000 field and laboratory measurements, Energy Build. 206 (2020) 109582, https://doi.org/10.1016/j.enbuild.2019.109582.

[50] A. Enander, Effects of moderate cold on performance of psychomotor and cognitive tasks, Ergonomics 30 (1987) 1431–1445, https://doi.org/10.1080/00140138708966037.

[51] L. Fang, D.P. Wyon, G. Clausen, P.O. Fanger, Impact of indoor air temperature and humidity in an office on perceived air quality, SBS symptoms and performance, Indoor Air 14 (2004) 74–81, https://doi.org/10.1111/j.1600-0668.2004.00276.x.

[52] V. Földváry Ličina, T. Cheung, H. Zhang, R. de Dear, T. Parkinson, E. Arens, C. Chun, S. Schiavon, M. Luo, G. Brager, P. Li, S. Kaam, M.A. Adebamowo, M. M. Andamon, F. Babich, C. Bouden, H. Bukovianska, C. Candido, B. Cao, S. Carlucci, D.K.W. Cheong, J.-H. Choi, M. Cook, P. Cropper, M. Deuble, S. Heidari, M. Indraganti, Q. Jin, H. Kim, J. Kim, K. Konis, M.K. Singh, A. Kwok, R. Lamberts, D. Loveday, J. Langevin, S. Manu, C. Moosmann, F. Nicol, R. Ooka, N.A. Oseland, L. Pagliano, D. Petráš, R. Rawal, R. Romero, H.B. Rijal, C. Sekhar, M. Schweiker, F. Tartarini, S. Tanabe, K.W. Tham, D. Teli, J. Toftum, L. Toledo, K. Tsuzuki, R. De Vecchi, A. Wagner, Z. Wang, H. Wallbaum, L. Webb, L. Yang, Y. Zhu, Y. Zhai, Y. Zhang, X. Zhou, Development of the ASHRAE global thermal comfort database II, Build. Environ. 142 (2018) 502–512, https://doi.org/10.1016/j.buildenv.2018.06.022.

[53] J.A. Porras-Salazar, S. Schiavon, P. Wargocki, T. Cheung, K.W. Tham, Indoor Temperature- Office Work Performance Database, Dryad, Dataset, 2021, doi.org/10.6078/D1G42R.

[54] O. Seppänen, W.J. Fisk, Q.H. Lei, Ventilation and performance in office work, Indoor Air 16 (2006) 28–36, https://doi.org/10.1111/j.1600-0668.2005.00394.x.

[55] P. Wargocki, J.A. Porras-Salazar, S. Contreras-Espinoza, W. Bahnfleth, The relationships between classroom air quality and children's performance in school, Build. Environ. 173 (2020) 106749, https://doi.org/10.1016/j.buildenv.2020.106749.

[56] J.A. Huwaldt, S. Steinhorst, Plot digitizer. http://plotdigitizer.sourceforge.net, 2015.

[57] G. James, D. Witten, T. Hastie, R. Tibshirani, An Introduction to Statistical Learning, Springer, 2013.

[58] R. R Core Team, A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2019. https://www.R-project.org/.

[59] M.C. Peel, B.L. Finlayson, T.A. McMahon, Updated world map of the Köppen-Geiger climate classification, Hydrol. Earth Syst. Sci. (2007) 12.

[60] D.W. Chiles, Effects of elevated temperatures on performance of a complex mental task, Ergonomics 2 (1958) 89–96, https://doi.org/10.1080/00140135808930404.

[61] R. Niemela, M. Hannula, S. Rautio, K. Reijula, J. Railio, The Effect of Air Temperature on Labour Productivity in Call Centres—A Case Study, vol. 6, Energy Build, 2002, https://doi.org/10.1016/S0378-7788(02)00094-4.

[62] K.W. Tham, H.C. Willem, Temperature and ventilation effects on performance and neurobehavioral-related symptoms of tropically acclimatized call center operators near thermal neutrality, Build. Eng. 111 (2005).

[63] R.D. Pepler, Warmth and performance: an investigation in the tropics, Ergonomics 2 (1958) 63–88, https://doi.org/10.1080/00140135808930403.

[64] N.H. Mackworth, Effects of heat on wireless operators, Br. J. Ind. Med. 3 (1946) 14316.

[65] W.H. Teichner, R.F. Wehrkamp, Visual-motor performance as a function of short-duration ambient temperature, J. Exp. Psychol. 47 (1954) 447–450, https://doi.org/10.1037/h0060272.

[66] B. Givoni, Y. Rim, Effect of the thermal environment and psychological factors upon subjects' responses and performance of mental work, Ergonomics 5 (1962) 99–114, https://doi.org/10.1080/00140136208930562.

[67] I.D. Griffiths, P.R. Boyce, Performance and thermal comfort, Ergonomics 14 (1971) 457–468, https://doi.org/10.1080/00140137108931266.

[68] M.A. Allen, G.J. Fischer, Ambient temperature effects on paired associate Learning*, Ergonomics 21 (1978) 95–101, https://doi.org/10.1080/00140137808931700.

[69] P.A. Bell, Effects of noise and heat stress on primary and subsidiary task performance, Hum. Factors 20 (1978) 749–752, https://doi.org/10.1177/001872087802000614.

[70] V.M. Sharma, M.R. Panwar, Variations in mental performance under moderate cold stress, Int. J. Biometeorol. 31 (1987) 85–91, https://doi.org/10.1007/BF02192842.

[71] I. Vasmatzidis, R.E. Schlegel, P.A. Hancock, An investigation of heat stress effects on time-sharing performance, Ergonomics 45 (2002) 218–239, https://doi.org/10.1080/00140130210121941.

[72] T. Witterseh, D.P. Wyon, G. Clausen, The effects of moderate heat stress and open-plan office noise distraction on SBS symptoms and on the performance of office work, Indoor Air 14 (2004) 30–40, https://doi.org/10.1111/j.1600-0668.2004.00305.x.

[73] L. Lan, Z. Lian, Use of neurobehavioral tests to evaluate the effects of indoor environment quality on productivity, Build. Environ. 44 (2009) 2208–2217, https://doi.org/10.1016/j.buildenv.2009.02.001.

[74] L. Lan, Z. Lian, L. Pan, Q. Ye, Neurobehavioral approach for evaluation of office workers' productivity: the effects of room temperature, Build. Environ. 44 (2009) 1578–1588, https://doi.org/10.1016/j.buildenv.2008.10.004.

[75] K.W. Tham, H.C. Willem, Room air temperature affects occupants' physiology, perceptions and mental alertness, Build. Environ. 45 (2010) 40–44, https://doi.org/10.1016/j.buildenv.2009.04.002.

[76] W. Cui, G. Cao, Q. Ouyang, Y. Zhu, Influence of dynamic environment with different airflows on human performance, Build. Environ. 62 (2013) 124–132, https://doi.org/10.1016/j.buildenv.2013.01.008.

[77] H. Maula, V. Hongisto, L. Östman, A. Haapakangas, H. Koskela, J. Hyönä, The effect of slightly warm temperature on work performance and comfort in open-plan offices - a laboratory study, Indoor Air 26 (2015) 286–297, https://doi.org/10.1111/ina.12209.

[78] B.M. Trezza, D. Apolinario, R.S. de Oliveira, A.L. Busse, F.L.T. Gonçalves, P.H. N. Saldiva, W. Jacob-Filho, Environmental heat exposure and cognitive performance in older adults: a controlled trial, Age 37 (2015) 43, https://doi.org/10.1007/s11357-015-9783-z.

[79] W. Liu, W. Zhong, P. Wargocki, Performance, acute health symptoms and physiological responses during exposure to high air temperature and carbon dioxide concentration, Build. Environ. 114 (2017) 96–105, https://doi.org/10.1016/j.buildenv.2016.12.020.

[80] S. Schiavon, B. Yang, Y. Donner, V.-C. Chang, W. Nazaroff, Thermal comfort, perceived air quality, and cognitive performance when personally controlled air movement is used by tropically acclimatized persons, Indoor Air 27 (2017) 690–702, https://doi.org/10.1111/ina.12352.

[81] F. Zhang, S. Haddad, B. Nakisa, M.N. Rastgoo, C. Candido, D. Tjondronegoro, R. de Dear, The effects of higher temperature setpoints during summer on office workers' cognitive load and thermal comfort, Build. Environ. 123 (2017) 176–188, https://doi.org/10.1016/j.buildenv.2017.06.048.

[82] X. Fan, W. Liu, P. Wargocki, Physiological and psychological reactions of sub-tropically acclimatized subjects exposed to different indoor temperatures at a relative humidity of 70%, Indoor Air 29 (2019) 215–230, https://doi.org/10.1111/ina.12523.

[83] A. Hedge, D. Gaygen, Indoor environment conditions and computer work in an office, HVAC R Res. 16 (2010) 123–138, https://doi.org/10.1080/10789669.2010.10390897.

[84] S. Hygge, I. Knez, Effects of noise, heat and indoor lighting on cognitive performance and self-reported affect, J. Environ. Psychol. 21 (2001) 291–299, https://doi.org/10.1006/jevp.2001.0222.

[85] L. Lan, Z. Lian, L. Pan, The effects of air temperature on office workers' well-being, workload and productivity-evaluated with subjective ratings, Appl. Ergon. 42 (2010) 29–36, https://doi.org/10.1016/j.apergo.2010.04.003.

[86] F. Obayashi, K. Miyagi, K. Ito, K. Taniguchi, H. Ishii, H. Shimoda, Objective and quantitative evaluation of intellectual productivity under control of room airflow, Build. Environ. 149 (2019) 48–57.

[87] T.Y. Chang, A. Kajackaite, Battle for the thermostat: gender and the effect of temperature on cognitive performance, PloS One 14 (2019), https://doi.org/10.1371/journal.pone.0216362 e0216362.

[88] P.A. Hancock, J.S. Warm, A dynamic model of stress and sustained attention, Hum. Factors J. Hum. Factors Ergon. Soc. 31 (1989) 519–537, https://doi.org/10.1177/001872088903100503.

[89] J.P. Campbell, B.M. Wiernik, The modeling and assessment of work performance, Annu. Rev. Organ. Psychol. Organ. Behav. 2 (2015) 47–74, https://doi.org/10.1146/annurev-orgpsych-032414-111427.

[90] Y.W. Ramírez, D.A. Nembhard, Measuring knowledge worker productivity: a taxonomy, J. Intellect. Cap. 5 (2004) 602–628, https://doi.org/10.1108/14691930410567040.

[91] T.H. Davenport, L. Prusak, Working Knowledge: How Organizations Manage what They Know, Harvard Business Press, 1998.

[92] D.P. Wyon, I. Wyon, F. Norin, Effects of moderate heat stress on driver vigilance in a moving vehicle, Ergonomics 39 (1996) 61–75, https://doi.org/10.1080/00140139608964434.

[93] L. Lan, L. Xia, R. Hejjo, D.P. Wyon, P. Wargocki, Perceived air quality and cognitive performance decrease at moderately raised indoor temperatures even when clothed for comfort, Indoor Air 30 (2020) 841–859, https://doi.org/10.1111/ina.12685.

[94] D.P. Wyon, P.O. Fanger, B.W. Olesen, C.J.K. Pederson, The mental performance of subjects clothed for comfort at two different air temperatures, Ergonomics 18 (1975) 359–374, https://doi.org/10.1080/00140137508931470.

[95] M. Schweiker, X. Fuchs, S. Becker, M. Shukuya, M. Dovjak, M. Hawighorst, J. Kolarik, Challenging the assumptions for thermal sensation scales, Build. Res. Inf. 45 (2017) 572–589, https://doi.org/10.1080/09613218.2016.1183185.