

Kramer, T. & Schiavon, S. (2026). Indoor temperature and office work performance in a temperate climate. *Buildings & Cities*.

SUPPLEMENTAL DATA

Instrumentation

A sensor mast located at the room center measured air temperature (-10 to 50°C , $\pm 0.1^{\circ}\text{C}$) and air speed (0.05 to 5.0 m/s, ± 0.02 m/s) at four heights (0.1 m, 0.6 m, 1.1 m, and 1.7 m) using the AirDistSys 5000 (Sensor Electronic, Poland, air velocity: 0.05 - 5 m/s, accuracy: ± 0.02 m/s $\pm 2\%$ of reading; temperature: -10 to $+50^{\circ}\text{C}$, accuracy: $\pm 0.2^{\circ}\text{C}$, sampling interval: 10 s). In addition, two Atmocube (Atmotech Inc., United States) IEQ monitoring units were placed at representative desk locations to measure air temperature (-40 to $+125^{\circ}\text{C}$, $\pm 0.5^{\circ}\text{C}$), relative humidity (0 to $100\%\text{RH}$, $\pm 2\%\text{RH}$), CO_2 (0 to 5000 ppm, ± 50 ppm $+ 2.5\%$ of reading at standard room levels), particulate matter (0 to 1000 $\mu\text{g}/\text{m}^3$, ± 5 $\mu\text{g}/\text{m}^3 + 5\%$ m.v. at 0 to 100 $\mu\text{g}/\text{m}^3$), light intensity (0 to 60000 lux, $\pm 10\%$), and background noise levels (30 to 100 dB, $\pm 2\text{dB}$) across the room.

PEBL Test Descriptions

In the Corsi Blocks test, a sequence of blocks lights up one at a time on screen and participants reproduce the sequence from memory, with sequences increasing in length until errors occur, measuring visuospatial working memory capacity. The Berg Card Sorting Test (BCST) requires participants to sort cards according to a rule that changes periodically without warning, measuring cognitive flexibility and set-shifting. The Flanker Task presents a central arrow flanked by congruent or incongruent arrows and asks participants to respond to the central arrow direction, measuring selective attention and inhibitory control. The Tower of London requires participants to rearrange colored beads into a target configuration in the minimum number of moves, measuring planning and problem-solving. The Trail Making test includes a simple condition (connect numbered dots sequentially, measuring processing speed) and a mixed condition (alternate between numbered and lettered dots, measuring set-shifting and executive function). The Balloon Analog Risk Task (BART) presents a simulated balloon that participants may pump for monetary rewards; they must decide when to stop before the balloon bursts, measuring risk-taking behavior and decision making. Full descriptions of the Paired Associates and Match-to-Sample tasks are provided in Porras-Salazar et al. (2024).

CAT learning effect

To examine whether participants improved with repeated exposure to the CAT, data were reorganized by trial number, independent of session temperature. **Figure S1** illustrates a clear upward trend across the four administrations of the test. Mean scores increased from 41.6 points on the first attempt ($\text{SD} = 15$) to 54.9 points by the fourth attempt ($\text{SD} = 14.7$). Intermediate scores followed the same trajectory, rising to 48.3 points at the second trial and 51.3 points at the third. This pattern indicates a consistent learning effect: participants became more familiar with the test format and content over time, which led to higher CAT scores regardless of the temperature condition under which the test was completed. Prior to the experiment, participants attended an introductory session during which they completed two condensed CAT practice test (6 minutes instead of 12 minutes) of comparable structure and difficulty to mitigate the learning effect during the experimental sessions. These practice sessions were anonymized and therefore not included in this analysis.

This clear presence of a learning effect in CAT performance reinforces the importance of balancing session order across participants. Because repeated exposure to the test was expected to produce performance improvements independent of temperature, the study was designed so that the four temperature conditions were evenly distributed across all possible session positions. Each participant received a unique session sequence, ensuring that no temperature was disproportionately associated with earlier or later test administrations. In addition, although each CAT administration used a different set of items, the tests followed the same structure, timing, and difficulty profile, reducing the risk that test version differences would influence performance. This approach allowed genuine temperature effects to be separated from performance gains driven purely by familiarity with the test.

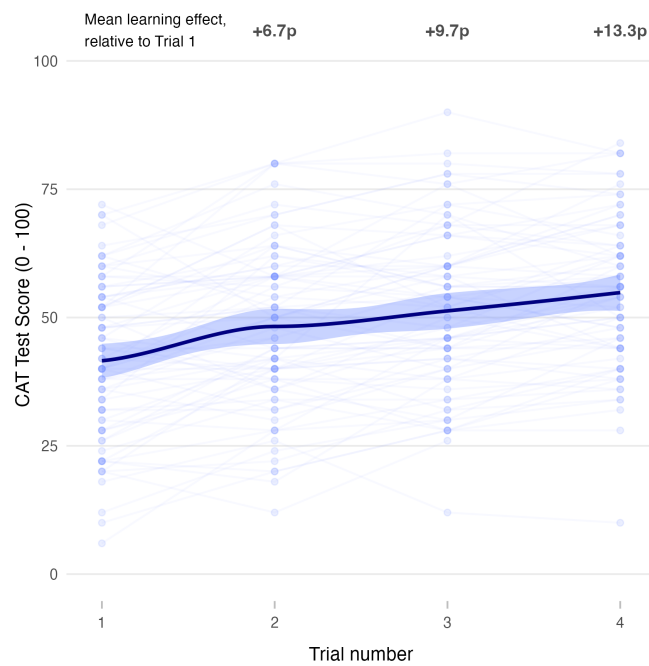


Figure S1: Learning effect in CAT. Changes in CAT scores across trials 1–4, where each point represents one test result, show a clear learning effect represented by higher scores in consecutive trials.

PEBL detailed results

Table S1 presents the complete test-by-test results for all PEBL metrics that were retained after excluding parameters with ceiling effects or negligible variation. For each metric, median values with interquartile ranges (IQR) are reported for each temperature condition, the percentage change relative to the 21°C baseline across the three warmer conditions, and repeated-measures ANOVA results testing for temperature effects. As shown in the table, no retained parameter exhibited statistically significant temperature effects (all $p > 0.11$ (70)). Performance changes were generally small (within ± 1 –5%) and showed no consistent directional pattern across related metrics or cognitive domains. These results indicate that cognitive performance across the tested domains (Working Memory, Executive Function, Attention, Processing Speed, and Decision Making) remained stable across the 21–27°C temperature range. Although some individual metrics showed slightly larger descriptive shifts at specific temperatures (e.g., attention-related metrics at 25°C), these changes were not statistically reliable and occurred alongside opposite-direction changes in related tasks, suggesting they reflect normal test-retest variability rather than systematic temperature effects.

Table S1: Detailed PEBL test battery results by temperature condition. For each retained metric, the table shows the median (IQR) performance at each temperature, percentage changes relative to 21°C baseline, and repeated-measures ANOVA results. Metrics are organized by cognitive domain and test. All p-values exceeded 0.05, indicating no statistically significant temperature effects on any measured cognitive parameter.

Test	Domain	Metric	Parameter	21°C	23°C	25°C	27°C	Performance change	p value	Effect size
BART	Decision making	Behavior	Mean Pumps	10.67 (6.61, 14.72, N = 76)	10.39 (4.86, 14.20, N = 78)	11.37 (6.56, 15.02, N = 76)	10.93 (5.19, 15.04, N = 77)	23C: -2.6%; 25C: +6.5%; 27C: +2.4%	0.87 (70)	0.003
BART	Decision making	Behavior	Total Pumps	958.00 (595.25, 1325.00, N = 76)	935.75 (437.00, 1278.25, N = 78)	1023.00 (590.50, 1352.50, N = 76)	983.50 (466.50, 1353.00, N = 77)	23C: -2.3%; 25C: +6.8%; 27C: +2.7%	0.90 (70)	0.003
BCST	Executive function	Speed	Mean Reaction Time	1046.38 (911.86, 182.50, N = 76)	1037.99 (915.97, 1239.85, N = 77)	1033.53 (887.27, 1222.12, N = 76)	1031.94 (923.04, 1218.89, N = 77)	23C: -0.8%; 25C: -1.2%; 27C: -1.4%	0.73 (69)	0.005
BCST	Executive function	Accuracy	Perseverative Errors	11.50 (10.00, 14.12, N = 76)	12.00 (10.50, 14.25, N = 77)	12.00 (10.50, 13.50, N = 76)	11.00 (10.50, 13.50, N = 77)	23C: +4.3%; 25C: +4.3%; 27C: -4.3%	0.36 (69)	0.015
BCST	Executive function	Efficiency	Number of Trials	120.00 (114.38, 128.00, N = 76)	121.00 (115.25, 125.50, N = 77)	121.00 (114.50, 128.00, N = 76)	120.00 (113.50, 126.50, N = 77)	23C: +0.8%; 25C: +0.8%; 27C: +0.0%	0.76 (69)	0.005
BCST	Executive function	Consistency	SD Reaction Time	487.19 (384.25, 704.27, N = 76)	507.88 (402.39, 743.93, N = 77)	496.72 (375.17, 759.58, N = 76)	492.89 (364.92, 706.84, N = 77)	23C: +4.2%; 25C: +2.0%; 27C: +1.2%	0.38 (69)	0.012
Corsi	Working memory	Accuracy	Mean Span	6.25 (5.50, 7.00, N = 76)	6.00 (5.50, 7.00, N = 78)	6.25 (5.75, 7.00, N = 76)	6.25 (5.50, 6.75, N = 77)	23C: -4.0%; 25C: +0.0%; 27C: +0.0%	0.69 (70)	0.007
Corsi	Working memory	Accuracy	Memory Span	75.00 (54.00, 96.00, N = 76)	72.50 (56.50, 96.00, N = 78)	78.00 (60.00, 96.00, N = 76)	75.00 (60.00, 94.00, N = 77)	23C: -3.3%; 25C: +4.0%; 27C: +0.0%	0.85 (70)	0.004
Corsi	Working memory	Accuracy	Total Correct Trials	10.50 (9.00, 12.00, N = 76)	10.00 (9.00, 12.00, N = 78)	10.50 (9.50, 12.00, N = 76)	10.50 (9.00, 11.50, N = 77)	23C: -4.8%; 25C: +0.0%; 27C: +0.0%	0.69 (70)	0.007
Flanker	Attention	Speed	Mean Response Time	408.83 (385.18, 444.29, N = 76)	415.04 (394.79, 448.31, N = 76)	408.33 (387.99, 439.06, N = 76)	415.70 (383.44, 438.61, N = 77)	23C: +1.5%; 25C: -0.1%; 27C: +1.7%	0.51 (68)	0.011
Flanker	Attention	Accuracy	Correct Responses	153.75 (149.88, 156.62, N = 76)	153.50 (149.62, 156.50, N = 76)	153.00 (151.00, 156.00, N = 76)	153.00 (149.50, 156.00, N = 77)	23C: -0.2%; 25C: 0.5%; 27C: -0.5%	0.73 (68)	0.006
Flanker	Attention	Accuracy	Errors	6.25 (3.38, 10.12, N = 76)	6.50 (3.50, 10.38, N = 76)	7.00 (4.00, 9.00, N = 76)	7.00 (4.00, 10.50, N = 77)	23C: +4.0%; 25C: +12.0%; 27C: +12.0%	0.73 (68)	0.006
Match To Sample	Attention	Speed	Mean Response Time	1121.44 (928.01, 1363.06, N = 76)	1153.58 (980.41, 1502.93, N = 78)	1167.97 (982.55, 1387.72, N = 76)	1181.57 (1010.50, 1426.68, N = 77)	23C: +2.9%; 25C: +4.1%; 27C: +5.4%	0.20 (70)	0.023
Match To Sample	Attention	Accuracy	Correct Responses	27.50 (25.50, 28.62, N = 76)	28.00 (26.50, 29.00, N = 78)	27.00 (26.00, 28.50, N = 76)	27.50 (25.50, 29.00, N = 77)	23C: +1.8%; 25C: -1.8%; 27C: +0.0%	0.68 (70)	0.005
Match To Sample	Attention	Accuracy	Errors	2.50 (1.00, 3.62, N = 76)	2.00 (1.00, 3.50, N = 78)	3.00 (1.50, 4.00, N = 76)	2.50 (1.00, 4.50, N = 77)	23C: -20.0%; 25C: +20.0%; 27C: +0.0%	0.50 (70)	0.011

Paired Assoc.	Attention	Speed	Mean Response Time	3221.85 (2746.24, 4195.60, N = 76)	3304.93 (2813.00, 4256.44, N = 78)	3175.17 (2607.81, 4104.60, N = 76)	3326.06 (2705.46, 4443.59, N = 77)	23C: +2.6%; 25C: -1.4%; 27C: +3.2%	0.83 (70)	0.004
Paired Assoc.	Attention	Accuracy	Correct Responses	22.00 (19.38, 23.50, N = 76)	22.50 (19.75, 24.00, N = 78)	22.50 (18.50, 23.50, N = 76)	22.50 (18.00, 24.00, N = 77)	23C: +2.3%; 25C: +2.3%; 27C: +2.3%	0.28 (70)	0.017
Paired Assoc.	Attention	Accuracy	Errors	2.00 (0.50, 5.12, N = 76)	1.50 (0.00, 4.25, N = 78)	1.50 (0.50, 5.50, N = 76)	1.50 (0.00, 6.00, N = 77)	23C: -25.0%; 25C: -25.0%; 27C: -25.0%	0.59 (70)	0.009
P-Trails	Executive function	Speed	Mixed: Total Time	29.75 (29.00, 32.00, N = 76)	30.00 (28.50, 32.00, N = 78)	29.50 (28.50, 32.00, N = 76)	30.50 (29.00, 32.00, N = 77)	23C: +0.8%; 25C: -0.8%; 27C: +2.5%	0.65 (70)	0.007
P-Trails	Processing speed	Speed	Letters: Total Time	31.00 (29.00, 33.50, N = 76)	32.00 (29.50, 34.62, N = 78)	32.00 (29.50, 35.50, N = 76)	31.50 (29.00, 35.00, N = 77)	23C: +3.2%; 25C: +3.2%; 27C: +1.6%	0.56 (70)	0.009
TOL	Executive function	Speed	Total Time	123.49 (88.83, 149.68, N = 76)	120.89 (96.83, 159.00, N = 78)	127.30 (93.85, 170.50, N = 76)	124.26 (92.12, 174.81, N = 77)	23C: -2.1%; 25C: +3.1%; 27C: +0.6%	0.11 (70)	0.029

Session Order Distribution

Table S2 shows the distribution of temperature conditions across session positions (1st through 4th) for the full cohort, the high-performer subgroup (top 25% by average CAT score), and the low-performer subgroup (bottom 25%). Each cell reports the number of participants who experienced a given temperature condition at a given position. The full cohort shows near-perfect balance by design. Because subgroup membership was not known before the experiment, session order could not be balanced within these groups, and some imbalance is visible in the subgroup rows. The LMM approach used for the subgroup analyses (trial number as a fixed covariate) statistically accounts for position-related learning effects regardless of their distribution across conditions.

Table S2: Session order distribution by group. Values show the number of participants who experienced each temperature condition at each session position (1st–4th), for the full cohort, high performers (top 25% by average CAT score), and low performers (bottom 25%).

Group	Temperature	Position 1	Position 2	Position 3	Position 4
Full cohort	21C	18	18	19	17
Full cohort	23C	21	19	19	16
Full cohort	25C	18	20	15	22
Full cohort	27C	18	18	22	17
High performers	21C	3	5	6	4
High performers	23C	6	3	5	4
High performers	25C	4	6	4	4
High performers	27C	5	4	3	6
Low performers	21C	4	6	3	3
Low performers	23C	5	4	6	3
Low performers	25C	5	6	2	5
Low performers	27C	4	2	7	5