

The Differentiated Effects Of Spatial Conditions In Educational Settings On Preference And EEG-Based Affect: A Comparative Study In Two Classrooms

Yavuz Bahadır Certel¹, Hacer Mutlu Danacı²

¹ Architecture Department, Faculty of Architecture, Akdeniz University, Antalya, Türkiye

² Architecture Department, Faculty of Architecture, Akdeniz University, Antalya, Türkiye

1st bahadircertel@akdeniz.edu.tr; 2nd hacermutlu@akdeniz.edu.tr

¹ <https://orcid.org/0000-0001-6800-4719>; ² <https://orcid.org/0000-0002-7325-6168>

*Hacer MUTLU DANACI: hacermutlu@akdeniz.edu.tr

Abstract: The physical characteristics of educational spaces can shape students' spatial preferences and the affective patterns that emerge during task performance. This study compared two classroom conditions (Classroom-1 vs Classroom-2) differing in a bundle of spatial attributes (light level, openness/window configuration, and volumetric features such as ceiling height), rather than isolating a single lighting effect. First, 128 students completed a preference survey. Next, 30 students drawn from the survey sample performed short exam-like tasks (mathematics and translation) in both classrooms using a counterbalanced within-subject design. EEG was recorded with Emotiv EPOC X. After quality control, valence-arousal (VA) analyses were conducted for $n = 29$, and band power/topography plus Emotiv Performance Metrics (PM) for $n = 24$. Matched comparisons used paired t /Wilcoxon tests, effect size (d_z), bootstrap 95% confidence intervals, and Benjamini-Hochberg FDR correction. Survey responses showed a clear preference for the brighter, more open condition. During task performance, EEG-based VA results indicated significantly higher valence in Classroom-2 ($t(28)=2.28$, $p=0.031$, $d_z=0.42$; 95% CI [0.06, 0.56]), with no significant change in arousal ($p=0.463$). Transition analyses showed no systematic VA category shifts between conditions. While most band-power indices did not differ, topographic analysis revealed a robust change in lateral frontal alpha asymmetry (F8-F7) after FDR ($q=0.028$). PM-Engagement decreased only before correction. These findings suggest that stated spatial preferences and task-related affect may diverge; combining surveys with neurophysiological measures can strengthen evidence for educational space design decisions.

Keywords—Educational spaces, EEG, neuroarchitecture, spatial conditions, visual comfort.

I. INTRODUCTION

Education plays a critical role in enabling individuals to develop their potential and, collectively, supporting social, cultural, and economic progress. Yet educational outcomes are not shaped by pedagogy and curriculum alone: the physical environment in which learning takes place contributes to students' engagement, wellbeing, and performance. School buildings and classrooms therefore function not merely as containers for instruction, but as active settings that can facilitate-or constrain-learning-related behaviours and cognitive-emotional states.

A substantial body of evidence suggests that school facility quality and building conditions are associated with student outcomes and teacher effectiveness. Early syntheses have reported measurable differences in achievement linked to facility condition, although the strength of evidence varies across studies and measurement approaches [1], [2]. Beyond overall building condition, classroom-level environmental factors such as thermal conditions, air quality, acoustics, furniture and layout, and aesthetic qualities can influence students' comfort, attention, and classroom behavior. Importantly, the educational environment is rarely experienced as isolated variables; interactions between environmental components can be as consequential as the components themselves [3].

Within this multi-factor context, light and daylighting have received particular attention because of their links to both visual comfort and non-visual biological pathways. Experimental and review evidence indicates that light exposure can affect alertness, subjective wellbeing, sleepiness, and aspects of cognitive performance, depending on characteristics such as intensity, duration, and spectrum [4], [5]. In educational settings, holistic field evidence also points to classroom design effects on learning progression, with the “Light” factor-constructed to reflect adequate daylight without glare and effective controllability-emerging as one of the influential classroom parameters [6]. However, even when daylighting is beneficial, the lived experience of “a bright classroom” is inseparable from other spatial properties that shape perception: window configuration, perceived openness and enclosure, surface reflectance, and volumetric characteristics can all modulate how light is distributed and how the space is interpreted.

This matter because students’ evaluations of a classroom (e.g., preference, perceived comfort, perceived attractiveness) do not necessarily track their cognitive-emotional state during a demanding task. The emerging neuroarchitecture literature argues that built environments can influence emotion and cognition through measurable neurophysiological correlates as well as self-report, yet robust controlled protocols remain limited and heterogeneous [7]. Specifically, for learning contexts, a recent systematic review synthesized evidence that built indoor features-including form/geometry, spatial distribution, dimensions such as height/width/enclosure, and light/sound/temperature-can be associated with attention and memory outcomes, while also highlighting methodological limitations and the need for more standardized approaches [8]. These reviews collectively suggest that focusing on a single environmental attribute (e.g., “daylight effect”) risks oversimplifying the broader perceptual and neurocognitive system through which learners respond to space.

A key component of that system is the volumetric character of space, particularly ceiling height and perceived enclosure. Experimental work in consumer and environmental psychology has proposed that salient ceiling height can prime concepts of freedom versus confinement, which in turn biases processing style toward more relational versus item-specific processing [9]. This priming framework is consistent with broader claims that spatial cues can modulate how information is approached and organized during cognition. In parallel, neuroaesthetic and neuroimaging evidence has shown that architectural features such as ceiling height and enclosure influence aesthetic judgments and approach-avoidance decisions, with corresponding neural activation patterns in regions supporting visuospatial exploration and affective appraisal [10]. Virtual-reality studies further indicate that geometric features of interior spaces relate systematically to affective appraisals [11]. Complementing these lines, EEG-based work has suggested that ceiling height can be associated with cognitive processing and wayfinding behavior in architectural settings [12]. Work in creativity research similarly discusses the role of spatial design in supporting creative thinking and comfort, aligning conceptually with the freedom/confinement account. Together, these strands suggest that “height and openness” are not merely aesthetic attributes; they can act as cognitive-emotional cues that shape appraisal and behavior.

These insights motivate a critical methodological point for educational-space research: real classrooms are experienced as bundled spatial conditions rather than isolated variables. Light level, openness/window configuration, and volumetric properties (including ceiling height) often co-vary in practice, and their combined effect may differ from what would be inferred from a single-factor study. Moreover, classrooms host different task types-learning activities, practice exercises, and exam-like assessments-that impose different cognitive demands and may evoke distinct affective patterns. Therefore, a design feature that is preferred subjectively (e.g., “brighter and more open”)

does not necessarily guarantee a consistent improvement in task-related affect or neurophysiological indicators during performance.

Against this background, the present study compares two classroom conditions (Classroom-1 and Classroom-2) that differ as integrated spatial configurations, including light level, openness/window configuration, and volumetric characteristics, rather than isolating a single lighting parameter. The study combines (i) a preference survey capturing students stated spatial preferences and (ii) a controlled within-subject experiment in which participants perform short exam-like tasks in both classrooms, while electroencephalography (EEG) is recorded. By analyzing valence–arousal indicators and complementary EEG-derived measures alongside self-reports, the study tests whether subjective preferences and task-related affective/neurophysiological responses converge or diverge when the spatial environment changes as a holistic package. In doing so, the work aims to contribute evidence relevant to educational design decisions: not only identifying which spatial condition is preferred, but also clarifying how bundled spatial properties relate to affective states during cognitively demanding classroom tasks.

II. MATERIALS AND METHODS

A. Data Collection Process

The study consisted of four stages: (i) on-site documentation and measurement of spatial characteristics in two classrooms, (ii) a preliminary preference survey, (iii) a controlled task-based experiment with EEG recording in both classrooms, and (iv) processing and analysis of EEG-derived metrics together with survey outputs. Spatial conditions were defined holistically as a bundle of co-varying features (light level, openness/window configuration, and volumetric characteristics such as ceiling height), rather than a single illuminance manipulation.

B. Experimental Design

Two classrooms representing different spatial conditions were used. In both spaces, students completed short exam-like performance tasks consisting of mathematics and language (translation) questions. To minimize order effects, classroom order was counterbalanced: half of the participants completed the tasks in the order Classroom-1 → Classroom-2, and the other half in the order Classroom-2 → Classroom-1. Different question sets were prepared by expert teachers with the goal of equivalent difficulty; sets were randomly distributed and applied in a balanced manner across conditions to reduce confounding between space and question-set effects.

Before starting, students were informed about the device and the aim of the experiment to reduce anxiety. After the headset was placed, a calibration period of approximately five minutes was used to ensure stable EEG signal reception. Based on pilot sessions indicating a decrease in engagement after ~5 minutes for tasks not affecting grades, each performance test was designed to last 2.5 minutes (150 s) per condition segment. Because the 2.5-minute segments were designed primarily to elicit comparable cognitive engagement rather than to estimate achievement, performance scores were not treated as a primary outcome; future work should include task accuracy and response-time measures. To reduce movement- and speech-related noise in EEG signals, participants were asked not to talk, avoid large muscle movements, and focus only on the questions. To reduce social distraction and maintain attention, each participant was tested alone in the classroom (Fig. 1).

Because students' mood may vary across the day, experiments were conducted after lunch between 2:30 p.m. and 4:30 p.m. All sessions took place in the same building under central heating and at similar times of day. Door/window status and seating arrangement

were kept standard to minimize external interference. Temperature, noise, and indoor air quality were not directly measured and are treated as limitations.

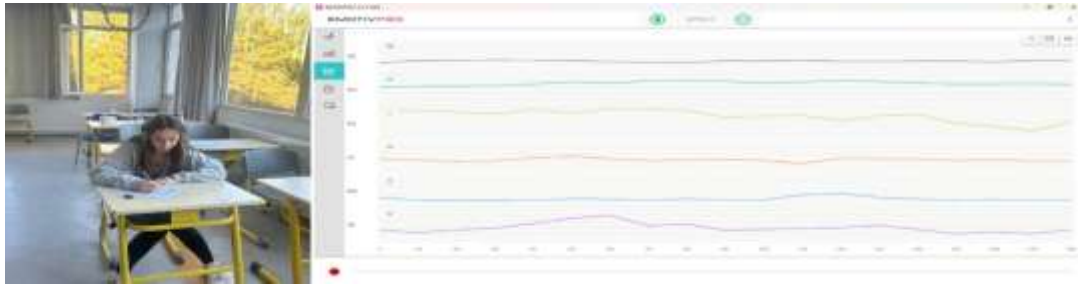


Fig. 1. A participant wearing the Emotiv EPOC X headset during the experiment and performance metrics

C. Tools and equipment

EEG was recorded using the Emotiv EPOC X mobile EEG system. The device provides 16 channels in total (14 EEG channels + 2 reference channels), transmits data via Bluetooth, and has a 14-bit resolution with a sampling rate of 128 Hz. (Fig. 2). The system was operated using a high-performance laptop connected to the internet (required because recordings are uploaded to the cloud system). Emotiv Builder was used to design the two-phase experiment and record data. Lux measurements were conducted using a Unit UT383S lux meter.



Fig. 2. Emotiv EPOC X mobile EEG headset and electrode placements, (Emotiv, 2025)

D. Experimental Settings

The experiments were conducted in two classrooms in a private college building in Antalya. The classrooms were selected to be similar in interior materials and colors, educational tools and equipment, furniture standards, and HVAC availability and typical operational settings (not instrumented), but to differ in ceiling height, openness to the outside, and window size/position-features that jointly shape daylight distribution and perceived visual comfort.

Classroom-1 measures 6.25×8.20 m (≈ 50 m²) and has a ceiling height of ≈ 2.90 m; daylight enters widely through large openings on two sides (Fig. 3A). Classroom-2 measures 7.20×6.10 m (≈ 43 m²) and has a ceiling height of ≈ 3.90 m; daylight enters through openings located on a single façade and at the upper level. (Fig. 3B). To quantitatively define the light-level component of the spatial conditions, illuminance measurements were taken on May 20, 2025 between 2:00 p.m. and 4:00 p.m. (the same window as the experiments), with artificial lighting turned off and curtains/blinds open. Measurements were taken at working-plane height (≈ 0.75 m) at 12 points using a 2×2 m

grid approach; values were reported as averages. Mean illuminance was 650 lux in Classroom-1 and 158 lux in Classroom-2.

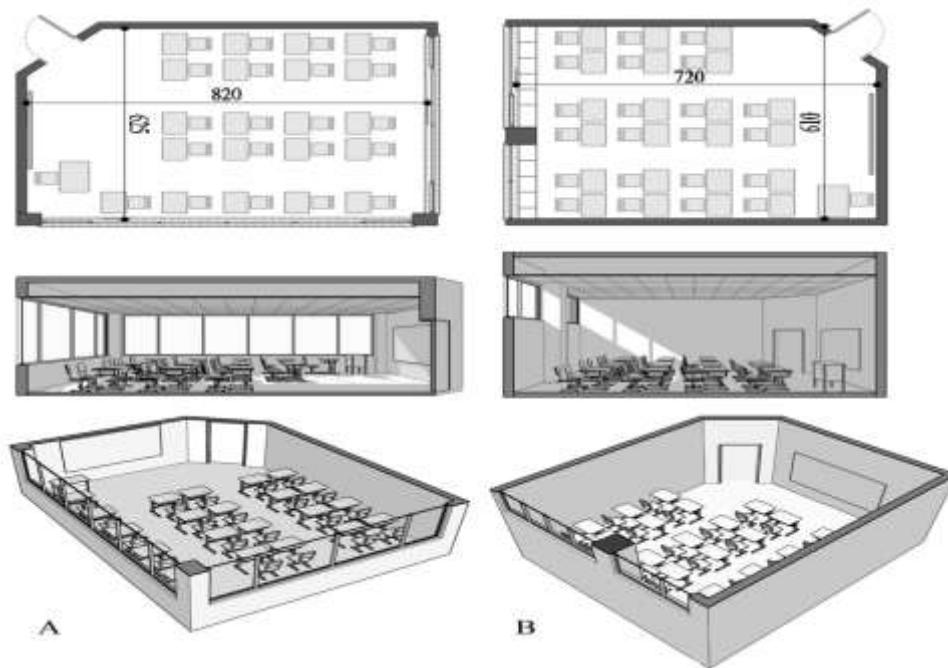


Fig. 3. Plan, section and perspective of Classroom-1 (A) and Classroom-2 (B)

Given that the classrooms differ simultaneously in light level (lux), exterior exposure/window configuration, and volumetric characteristics (ceiling height), the independent variable was treated as the overall spatial condition created by these physical characteristics together, and findings were interpreted within this holistic framework.

E. Participants

The participants in the experiment were selected from 10th grade students at the same college. Care was taken to ensure that the participants did not have any neurological or psychological illnesses or a history of illness and had no concerns about the experiment. Participants were selected on a voluntary basis. Furthermore, to avoid confounding factors such as participants feeling a sense of belonging to the space or having memories associated with it, care was taken to ensure that participants were not regular users of the classrooms at any point in time. The participants in the experiment consisted of 16 male and 14 female students aged 15-16.

Ethics Committee Approval: This study received ethical approval from the Ethics Committee of the relevant university with decision number 1201565 dated 07.05.2025. Data collection permissions were obtained by obtaining the necessary approvals from the administration of the school where the research was conducted and from the parents of the students participating in the research.

F. Collection and analysis of EEG data

EEG data were collected using the Emotiv EPOC X system. The process of obtaining emotion-related metrics from EEG consists of the following steps: (i) preprocessing, (ii) power spectrum and band power extraction, (iii) calculation of valence-arousal (VA)

indices, (iv) topographic visualization, (v) VA transition analysis, and (vi) extraction and evaluation of Emotiv Performance Metrics (PM) metrics (Fig. 4).

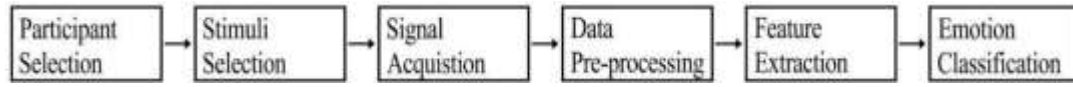


Fig. 4. Emotion extraction workflow from EEG signals.

Analyses were performed in a reproducible manner in a Python/Jupyter environment; numerical operations, statistics, tables, and figure generation were scripted. VA analyses after quality control were conducted for $n=29$. Band power/topographic map and PM analyses were reported for $n=24$ on the subsample meeting marker/IsActive conditions in the Emotiv PM stream; missingness was handled listwise on an analysis-by-analysis basis. EEG recordings followed the Emotiv channel set (AF3, AF4, F3, F4, F7, F8, FC5, FC6, T7, T8, P7, P8, O1, O2), corresponding to the international 10–20 system.

Preprocessing (Emotiv Analyzer)

Preprocessing was performed in Emotiv Analyzer using the following parameters: band-pass filtering 0.5–45 Hz; mean reference; slew limit 30; epoching with a 512-sample window and 256-sample step (overlapping segmentation); and band ranges of 1–4, 4–8, 8–13, 13–30, and 30–45 Hz. These steps were applied to reduce low-frequency drift and high-frequency noise, reduce common-mode components via referencing, and create fixed-length segments for feature extraction.

Power spectrum and bandpower

Band power metrics were obtained from the power spectrum of preprocessed EEG data. Analyses focused on commonly reported bands (particularly alpha and beta) and on frontal channels/pairs frequently associated with task-related state changes.

Valence-Arousal indices

Valence index: Valence was calculated based on the frontal alpha asymmetry approach. First, raw asymmetry Right-Left was defined as $FAA(\alpha) = \alpha(F4) - \alpha(F3)$ (8–13 Hz). Based on the assumption that alpha power is inversely related to cortical activity (high alpha $\rightarrow$ relatively low cortical activation), and consistent with the common interpretation in the literature of "relative left activity based on the right-left alpha difference" [13], [14], the valence axis is coded as $Valence = -FAA(F4-F3)$ with sign reversal for interpretability in reporting. Therefore, the sign of the valence score is reversed relative to the raw asymmetry; all interpretations follow this convention [15]

Arousal index: The arousal dimension has been operationalized in line with the family of ratio-based indices based on the relative dominance of the beta and alpha bands, which are frequently used in the EEG literature as indicators of arousal/engagement [16], [17], [18]. Based on the assumption that beta activity is associated with more alert/active mental states, while alpha activity is associated with more relaxed/low activation states [18], this study uses band power difference instead of ratio, as it carries the same directional information and is a more stable parameterization in terms of scaling/normalization: $Arousal = \text{mean_frontal}[\beta(13-30)] - \text{mean_frontal}[\alpha(8-13)]$. The index was calculated over a frontal channel pool (AF3, F7, F3, FC5, FC6, F4, F8, AF4). Since limited standardization and different formulations have been reported in EEG-derived

"involvement/engagement/arousal" indices regarding band and montage definitions [19], this study interprets the selected difference index within that family. VA scores were visualized on Russell's two-dimensional affect space [20]. To bring the axes in the figures to a comparable range, valence and arousal values were scaled using maximum absolute (max-abs) scaling ($x_{\text{scaled}} = x / \max(|x|)$) solely for visualization purposes; which scales

the values to the range [-1, 1] while preserving the sign [21]. The purpose of this presentation choice is to ensure compatibility and visual comparability with the [-1, +1] range commonly used in valence-arousal studies [22]. Statistical tests were performed on raw VA indices rather than scaled values.

Statistical evaluation (VA-continuous). Between-conditions differences for valence and arousal were tested using a matched-design logic. The normality assumption was assessed using the Shapiro-Wilk test on between-conditions difference scores [23]; when appropriate, a paired t-test was reported, and when the assumption was weak, the Wilcoxon signed-rank test was reported as a robustness check [24]. Effect size was reported using Cohen's d_z , appropriate for a matched design [25]. Hotelling's T^2 test was used to evaluate valence and arousal differences together in two dimensions [26]. 95% confidence intervals for mean differences were calculated using bootstrap resampling ($B = 10,000$; percentile).

Topographic maps

Topographic maps were generated using the average band power of all channels and reported in three panels: 1) Classroom-1, 2) Classroom-2, 3) $\Delta = \text{Classroom-2} - \text{Classroom-1}$.

Valence topographic map: alpha hemispheric asymmetry (Right-Left) was calculated over symmetric pairs: AF4-AF3, F4-F3, F8-F7, FC6-FC5, T8-T7, P8-P7, O2-O1.

Arousal topographic map: β - α index was calculated for each channel and displayed topographically. Statistical evaluation (topographic map/channel level). To limit false-positive accumulation in analyses involving multiple channel-based tests, the Benjamini-Hochberg FDR correction was used, and corrected results were reported [27].

VA transition analysis

To examine categorical shifts in VA space between conditions, transition analyses were conducted for binary summaries (valence sign; arousal activation) and 4-state/quadrant classifications, with agreement reported using Cohen's κ .

Emotiv Performance Metrics (PM)

Emotiv's derived PM scores (e.g., attention, engagement, relaxation, stress, etc.) were obtained from EmotivPRO outputs. PM scores were obtained from the PM.<metric>.Scaled fields in the pm.bp.csv files. To ensure data integrity, analysis windows were defined by considering the IsActive variable in the session stream, and the average PM value at the participant level was calculated for each phase. Due to manufacturer-specific calculation details, these metrics were interpreted as auxiliary indicators supporting EEG-based findings rather than as "precise emotion measurements." Statistical evaluation (PM). Differences between conditions were tested under a matched design using phase averages. Since multiple PM metrics were evaluated together, the Benjamini-Hochberg FDR correction was used for significance assessment [27].

Statistical analysis

Since both conditions were measured in the same participants in this study, the basic inference framework is a matched (within-subject) design. Statistical analyses were performed using a Python-based workflow (in a Jupyter Notebook environment); NumPy/Pandas, SciPy, and relevant visualization/EEG libraries were primarily used for data processing and testing. Preprocessing of EEG data and feature extraction were performed using Emotiv Analyzer outputs, and analysis files were then evaluated in a reproducible manner in the Python environment.

The paired t-test tests whether the mean difference between two conditions is significantly different from zero; one of the assumptions of this test is that the difference scores are approximately normally distributed. The normality assumption was therefore assessed using the Shapiro-Wilk test on the difference scores rather than the raw scores

[23]. In cases where the normality assumption was weak or sensitivity to outliers was considered possible, the Wilcoxon signed-rank test, a ranking-based and more robust alternative, was reported for robustness assessment [24]. Cohen's d_z , appropriate for matched designs, was used to report the effect size alongside "significance" [25]. When two variables, such as valence and arousal, needed to be evaluated together, Hotelling's T^2 test, which tests the simultaneous difference in two dimensions, was applied [26].

The Benjamini-Hochberg FDR approach was preferred to limit false-positive accumulation when conducting simultaneous hypothesis tests on multiple channels/pairs or multiple PM metrics [27]. FDR adjustment was applied separately within each family of analyses (e.g., homologous channel pairs/topographic comparisons; multiple PM metrics; band power measures), and adjusted q-values were reported. Stuart-Maxwell (marginal homogeneity), Bowker (symmetry), and McNemar (2×2 matched ratio difference) tests, along with Cohen's κ (chance-corrected agreement), were used in categorical transition analyses [28], [29], [30], [31], [32].

All tests were two-tailed; the significance level was set at $\alpha = 0.05$ (evaluation was based on q values reported in FDR adjustment). 95% confidence intervals for mean differences were calculated using bootstrap resampling (e.g., 10,000 repetitions; percentile) [33].

III. RESULTS AND DISCUSSIONS

A. Preliminary research on students' space preferences

In the first stage of the research, a survey was conducted to understand students' space preferences and reasons for their preferences. Students overwhelmingly preferred Classroom 1 for all activities, with a rate of 93%. Students' space preferences vary for different activities and encompass different reasons. The percentages of Classroom-1 and Classroom-2 preferences for exams, lessons, individual work, and group work in the survey involving 128 students are shown in Fig. 5.

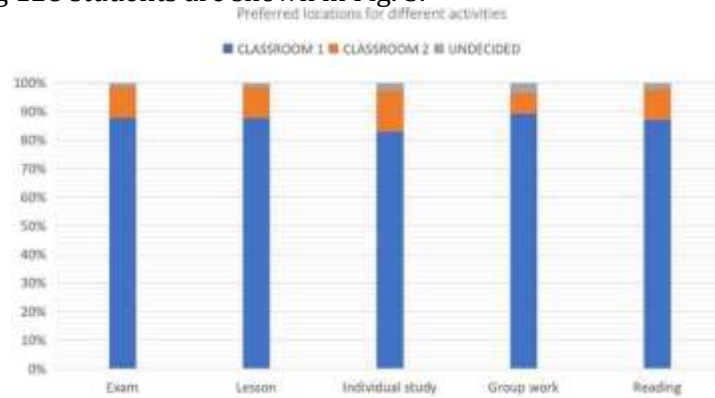


Fig. 5. Percentage distribution of students' space preferences for different activities

Students predominantly preferred Classroom 1 for all activities. They preferred Classroom 2 at the highest rate for individual study and at the lowest rate for group work. The students who participated in the survey stated that the reasons influencing their preferences were lighting, spaciousness and airiness, feeling comfortable, better focus, and air quality. Among these reasons for preference, lighting features ranked first, while ventilation features ranked last (Fig. 6).

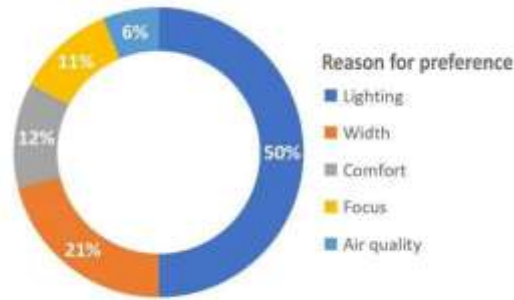


Fig. 6. Reasons for space preference

B. Valence/Arousal analyses and positioning in Russell space

EEG recordings taken during the mathematics and English-Turkish translation tests were analyzed and compared. The scores obtained from the tests were not evaluated, as it was considered that the scores obtained from a 2.5-minute performance test could not indicate the level of success. The main objective here is to comparatively examine EEG-based affect indicators (VA and complementary measures) that emerge during exam-like tasks in two different spatial conditions. Band power summaries and topographic visualizations of the EEG recordings obtained during the applied tests are presented in the relevant Figures (Figs. 8-13), while VA statistical summaries are presented in Table 1. The graphs of the measurements are labeled according to phases as classroom 1 and classroom 2. The affective positions of the participants in the Classroom-1 and Classroom-2 conditions were evaluated on Russell's two-dimensional valence-arousal (VA) space (Russell, 1980). For visualization purposes, the VA axes were scaled to the range $-1 \dots +1$ (max-abs; $x_{scaled} = x / \max(|x|)$), and statistical tests were performed on the raw VA indices (Fig. 7).

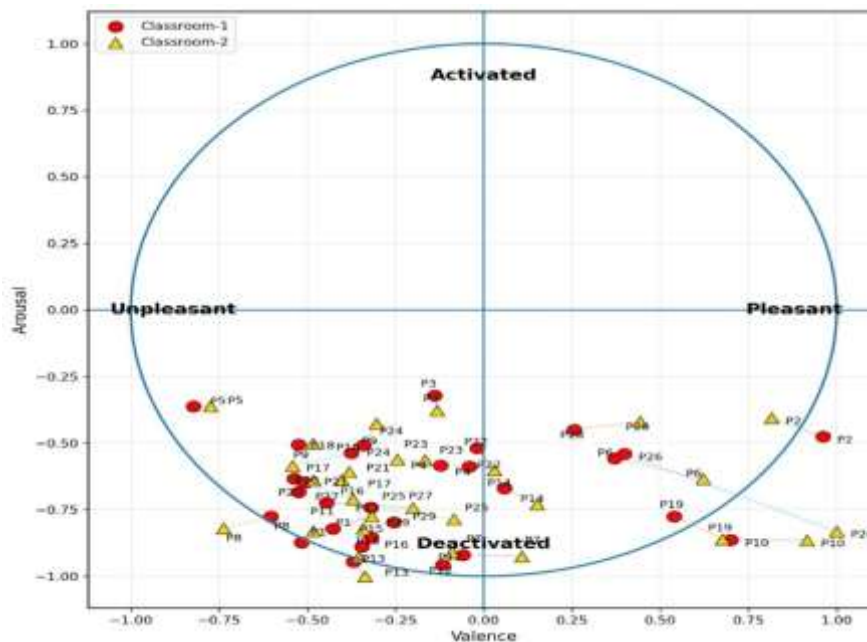


Fig. 7. Russell VA space (Classroom-1: red circle; Classroom-2: yellow triangle; participants P1...P29)

When examining the Russell space, a general shift toward a more positive direction (right) on the valence axis was observed in the Classroom-2 condition (Fig. 7). In the quantitative analysis, the valence mean was found to be $M = -0.67$ ($SD = 1.86$) in Classroom-1 and $M = -0.37$ ($SD = 2.14$) in Classroom-2. In the paired comparison, the valence difference is significant: $t(28) = 2.28$, $p = 0.031$, $d_{z} = 0.42$; bootstrap 95% CI [0.06, 0.56]. Due to deviation from the normality assumption of the distribution (Shapiro-Wilk $p = 0.018$), the Wilcoxon signed-rank test, applied as a robustness check, also supported the result ($W = 115$, $p = 0.026$) (Table 1) [24], [25].

No significant change was detected between conditions in the arousal dimension (Fig. 7). The arousal mean was $M = -4.85$ ($SD = 1.29$) in Classroom-1 and $M = -4.94$ ($SD = 1.32$) in Classroom-2, and the paired comparison was not significant: $t(28) = -0.74$, $p = 0.463$, $d_{z} = -0.14$; bootstrap 95% CI [-0.31, 0.13]. Due to the normality deviation (Shapiro-Wilk $p = 0.008$), the Wilcoxon test applied is also not significant ($W = 174$, $p = 0.358$) (Table 1).

TABLE I. PAIRED COMPARISONS FOR VALENCE AND AROUSAL IN CLASSROOM-1 AND CLASSROOM-2 CONDITIONS

Measure	n	Classroom-1 M (SD)	Classroom-2 M (SD)	Δ (C2-C1)	95% CI bootstrap	T (df)	P(t)	P Wilcoxon	dz
Valence	29	-0.67 (1.86)	-0.37 (2.14)	0.30	[0.06, 0.56]	2.28 (28)	0.031	0.026	0.42
Arousal	29	-4.85 (1.29)	-4.94 (1.32)	-0.08	[-0.31, 0.13]	-0.74 (28)	0.463	0.358	-0.14

When valence and arousal differences were evaluated together, Hotelling's T^2 test did not reach the significance level ($T^2 = 5.62$, $F(2, 27) = 2.71$, $p = 0.085$). This pattern indicates that the fundamental differentiation between conditions emerged on the valence axis rather than the arousal axis (Table 1) [26].

C. VA transition analysis: quadrant and 4-state

To examine potential "category changes" in the VA space, a two-level transition analysis was conducted: (i) binary summaries based on valence sign (Pleasant vs. Unpleasant), (ii) binary summaries based on arousal (Activated vs. Deactivated), and their combination (iii) 4-state/quadrant transitions. The Stuart-Maxwell test for marginal homogeneity (Stuart, 1955; Maxwell, 1970) and the Bowker symmetry test (Bowker, 1948) were used for multi-category matched tables; the McNemar test was applied for binary summaries [31]. Inter-condition classification agreement was reported using Cohen's κ (Table 2; McNemar-Valence and McNemar-Arousal rows).

In 4-state transitions, no evidence was found that marginal distributions changed: Stuart-Maxwell $\chi^2(3) = 3.71$, $p = 0.295$. No evidence was found that transitions had a directional/asymmetric structure: Bowker $\chi^2(6) = 4.33$, $p = 0.632$. The McNemar test is not significant for the valence sign in binary summaries ($p = 0.480$; mismatched pairs $b_{01}=2$, $b_{10}=0$); there is also no significant difference for the arousal binary classification ($p = 1.000$; $b_{01}=2$, $b_{10}=3$). In contrast, inter-rater classification agreement is moderate-to-high (4-state $\kappa = 0.657$).

Similarly, at the quadrant level, marginal homogeneity and symmetry tests are not significant (Stuart-Maxwell $\chi^2(3)=4.18$, $p = 0.243$; Bowker $\chi^2(6)=4.33$, $p = 0.632$). Quadrant agreement is $\kappa = 0.675$, indicating that the classification remains stable across conditions for a significant portion of participants (Table 2).

TABLE II. MARGINAL HOMOGENEITY, SYMMETRY, PAIRWISE SUMMARY TESTS, AND CROSS-CONDITION AGREEMENT FOR VA CATEGORICAL TRANSITIONS.

Analysis	Test	df	p	Notes
4-state	Stuart-Maxwell	3	0.295	
4-state	Bowker	6	0.632	
4-state	McNemar (Pleasant vs Unpleasant)	1	0.480	b01=2, b10= 0
4-state	McNemar (Activated vs. Deactivated)	1	1.000	b01=2, b10= 3
4-state	Cohen's kappa		0.657	
Quadrant	Stuart-Maxwell	3	0.243	
Quadrant	Bowker	6	0.632	
Quadrant	McNemar (Pleasant vs Unpleasant)	1	1.000	b01=2, b10= 1
Quadrant	McNemar (Activated vs. Deactivated)	1	1,000	b01=2, b10= 3
Quadrant	Cohen's kappa		0.675	

D. Band power analyses

To support valence and arousal findings at the EEG level, alpha (8-13 Hz) and beta (13-30 Hz) band powers (dB) of the F3 and F4 channels in the frontal region were compared. Additionally, frontal alpha asymmetry (FAA; F4-F3), which is assumed to be related to valence, and the β - α index derived for arousal [mean_frontal(β) - mean_frontal(α); AF3, F7, F3, FC5, FC6, F4, F8, AF4] were calculated (Coan & Allen, 2004; Davidson, 1992; Klimesch, 1999). Participant-based matched patterns are shown in Fig. 8, the time-series group summary in Fig. 9, and the statistical summary in Table 3. No significant change in band power was detected between conditions (all p's > 0.22; Wilcoxon p's > 0.16; Table 3). The F3 alpha power was -8.02 ± 2.32 dB in Classroom-1 and -7.76 ± 2.77 dB in Classroom-2, with no significant difference: $t(23) = 1.25$, $p = 0.223$, $d_{z} = 0.26$ (Wilcoxon $p = 0.208$). The difference for F4 alpha power is also not significant ($p = 0.723$). Although there is a slight upward trend for F3 in the beta band, it is not statistically significant ($p = 0.389$); the F4 beta band has remained practically unchanged ($p = 0.986$). Although the valence support measure FAA(α) appears to have decreased slightly in Classroom-2, this difference is also not significant ($t(23) = -0.66$, $p = 0.514$). The β - α index for arousal also did not show a significant change between conditions ($t(23) = -0.80$, $p = 0.432$). After applying multiple comparison correction (BH-FDR; $m=6$), $q > 0.05$ for all comparisons, thus all results remain non-significant. (Table 3).

TABLE III. COMPARISON OF FRONTAL F3/F4 ALPHA-BETA BAND POWERS AND DERIVED VALENCE/AROUSAL INDICES ACROSS CONDITIONS

Metric	C1 (M $\pm$ S D, dB)	C2 (M $\pm$ S D, dB)	Δ C2-C1 M (dB)	95% CI for Δ	t(df)	p (t)	d _z	Wil. W	p W	q_FDR
F3 α (8-13 Hz)	-8.02 $\pm$ 2.32	-7.76 $\pm$ 2.77	0.26	[-0.17, 0.69]	1.25 (23)	0.22 3	0.2 6	105	0.208	0.771

F4 α (8-13 Hz)	-7.14 ± 3.19	-7.07 ± 3.27	0.07	[-0.34, 0.49]	0.36 (23)	0.72 3	0.0 7	117	0.360	0.868
F3 β (13- 30 Hz)	-12.94 ± 1.71	-12.85 ± 1.63	0.09	[-0.12, 0.30]	0.88 (23)	0.38 9	0.1 8	101	0.169	0.771
F4 β (13- 30 Hz)	-11.97 ± 1.96	-11.96 ± 1.94	0.00	[-0.24, 0.24]	0.02 (23)	0.98 6	0.0 0	133	0.643	0.986
Valen ce FAA (α): F4 - F3	0.88 $\pm$ 1.92	0.69 $\pm$ 2.13	-0.19	[-0.77, 0.40]	-0.66 (23)	0.51 4	- 0.1 4	123	0.456	0.771
Arous al mean (β)- (α)	-4.88 ± 1.74	-5.00 ± 1.77	-0.12	[-0.43, 0.19]	-0.80 (23)	0.43 2	- 0.1 6	117	0.36	0.771

In the participant-based paired graph (Fig. 8), there is significant heterogeneity among individuals; some participants show increases in certain bands, while others show decreases. This suggests that the increase in valence in the VA may emerge more through individual patterns rather than as a unidirectional, strong group effect at the EEG band power level. The overall profiles of the two conditions largely overlap in the time series summaries; however, the sharp increases observed towards the end of the marker should be interpreted with caution as they may be sensitive to windowing/segment boundary (edge) effects (Fig. 9).

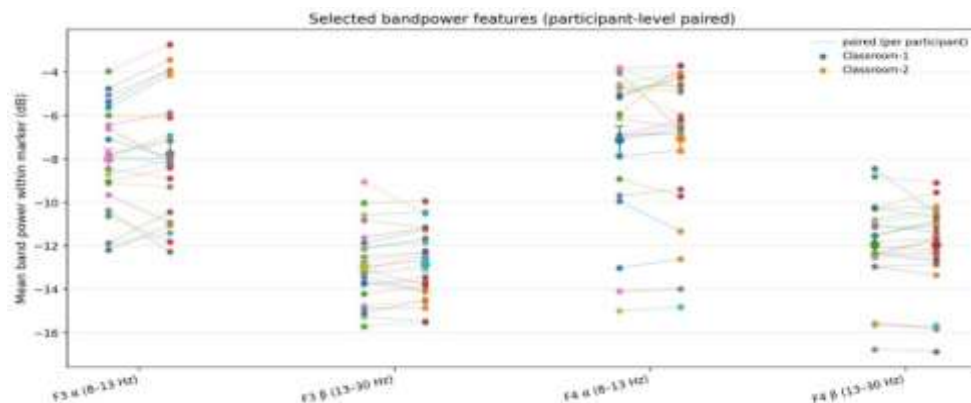


Fig. 8. Participant-based matched comparison for selected bandpower features (F3/F4; alpha/beta).

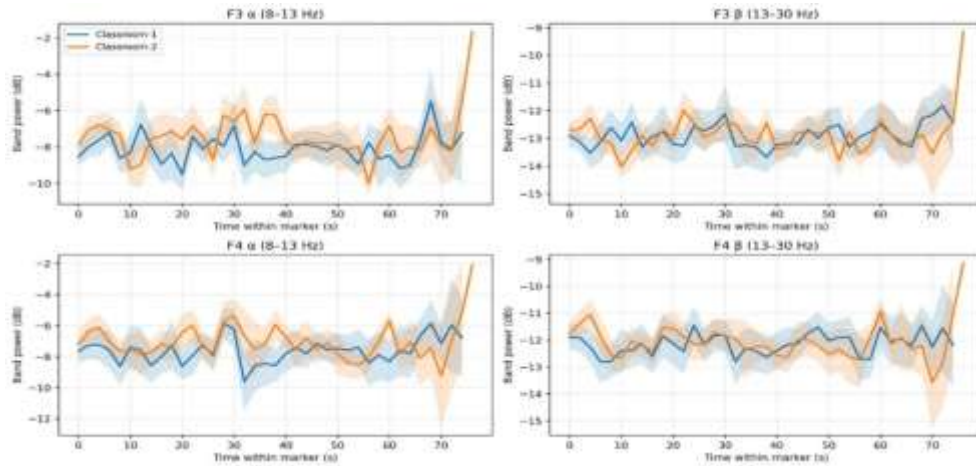


Fig. 9. Time-series group summary for F3/F4 band powers (alpha/beta).

E. Topographic maps with valence and arousal support

To examine valence and arousal findings topographically at the EEG level, topographic maps were created for (i) alpha band power (8-13 Hz), (ii) alpha hemispheric asymmetry (Right-Left) for valence support, (iii) the β - α index for arousal support, and (iv) the β/α ratio [34], [35], [36]. Panels for Classroom-1, Classroom-2, and the difference map ($\Delta = C2-C1$) were reported for each metric (Figures 11-14).

Alpha band power (8-13 Hz)

Alpha band power topography generally shows a similar distribution in both conditions (Fig. 10A-B). When examining the difference map ($\Delta = C2-C1$), it is observed that the magnitude of change is limited (approximately ± 0.36 dB scale) and increases remain at a "low-medium" level spread over a wider area (Fig. 10C; Table 4). This pattern suggests that alpha power does not show a topographically distinct and localized strong differentiation between conditions.

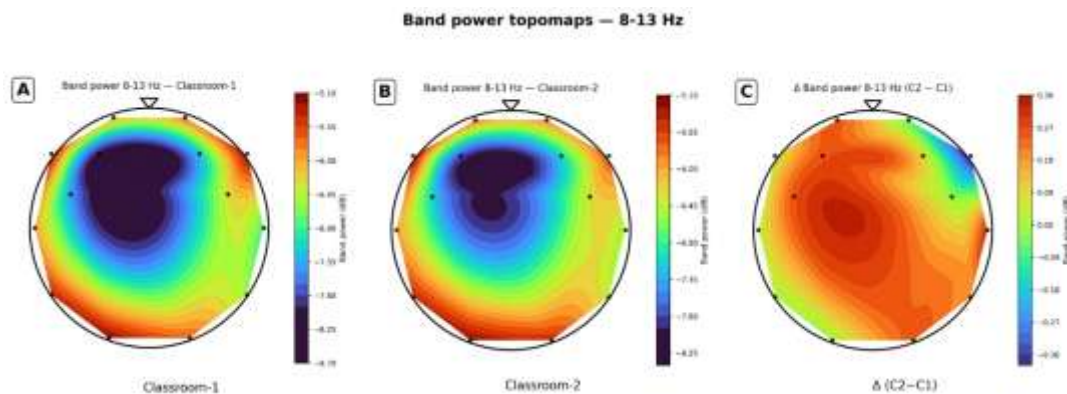


Fig. 10. Alpha (8-13 Hz) band power topographic maps (C1, C2, Δ).

TABLE IV. CHANNEL-LEVEL MATCHED STATISTICS FOR THE ALPHA (8-13 Hz) BAND POWER TOPOGRAPHIC MAP ($\Delta = C2-C1$).

Channel	n	Δ (C2-C1) mean	t	p	q_FDR	d_z	Sig (q<.05)
F3	24	0.299	2.024	0.055	0.308	0.413	FALSE

T8	24	0.324	1.988	0.059	0.308	0.406	FALSE
F8	24	-0.307	-1.927	0.066	0.308	-0.393	FALSE
P8	24	0.206	1.550	0.135	0.456	0.316	FALSE
O2	24	0.221	1.440	0.163	0.456	0.294	FALSE
AF3	24	0.178	1.164	0.256	0.597	0.238	FALSE
FC5	24	0.081	0.660	0.516	0.935	0.135	FALSE
O1	24	0.104	0.631	0.534	0.935	0.129	FALSE
F4	24	0.063	0.523	0.606	0.943	0.107	FALSE
F7	24	-0.047	-0.297	0.769	0.985	-0.061	FALSE
AF4	24	-0.037	-0.291	0.774	0.985	-0.059	FALSE
T7	24	0.015	0.103	0.919	0.997	0.021	FALSE
FC6	24	-0.007	-0.063	0.950	0.997	-0.013	FALSE
P7	24	0.000	0.003	0.997	0.997	0.001	FALSE

Valence support: alpha hemispheric asymmetry (Right-Left)

Hemispheric asymmetry maps assumed to be related to valence show a consistent distribution in both conditions (Fig. 11A-B). However, matched statistics across homologous channel pairs (FDR-corrected) revealed a significant change only for the F8-F7 pair (qFDR= 0.028; Δ mean = -0.45 dB; $t(23)=-3.19$, $p= 0.004$; $d_{z}=-0.65$). This result indicates that the right-left alpha asymmetry pattern (according to the Right-Left definition) in the lateral frontal region (F8-F7) changed significantly in the Classroom-2 condition. Although changes were observed for other frontal pairs (e.g., F4-F3, AF4-AF3), they did not reach the level of significance after multiple comparisons (Fig. 11C; Table 5).

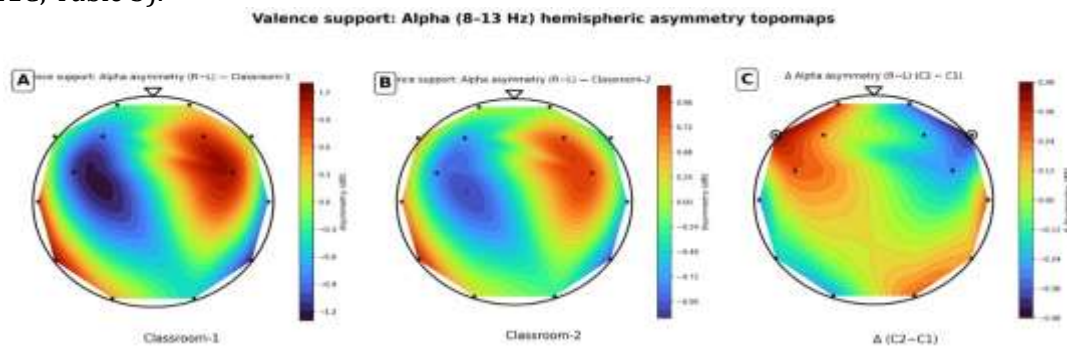


Fig. 11. Valence support: alpha hemispheric asymmetry (R-L) topographic maps (C1, C2, Δ).

TABLE V. VALENCE SUPPORT: HOMOLOGOUS CHANNEL PAIR COMPARISONS FOR ALPHA HEMISPHERIC ASYMMETRY (R-L) ($\Delta=C2-C1$).

Channel pair	n	Δ (C2-C1) mean	t	p	q_FDR	d_z	Sig (q<.05)
F8-F7	24	-0.451	-3.195	0.004	0.028	-0.652	TRUE
AF4-AF3	24	-0.332	-2.124	0.045	0.156	-0.434	FALSE
FC6-FC5	24	-0.271	-1.883	0.072	0.169	-0.384	FALSE
O2-O1	24	0.265	1.412	0.171	0.300	0.288	FALSE
T8-T7	24	0.298	1.132	0.269	0.355	0.231	FALSE
P8-P7	24	0.177	1.050	0.305	0.355	0.214	FALSE
F4-F3	24	-0.187	-0.663	0.514	0.514	-0.135	FALSE

This finding suggests that the valence increase observed in Russell space is not strongly captured by the classic F4–F3 asymmetry alone; instead, a more sensitive asymmetry change emerges in the lateral frontal region (F8–F7).

Arousal support (β - α index)

In the β - α index maps used for arousal, the overall topographic pattern across conditions is largely preserved (Fig. 12A-B). The difference map ($\Delta=C2-C1$) mostly indicates a negative pattern (index decrease) (Fig. 12C). Although some channels showed changes at the nominal level in channel-based paired tests (e.g., $p=0.034$ for O2), no channel was significant after FDR correction ($q < \text{FDR} > 0.05$). This result, consistent with the absence of significant changes in the arousal dimension in the VA analysis, suggests that the arousal support index also did not show a clear differentiation at the group level (Fig. 12; Table 6).

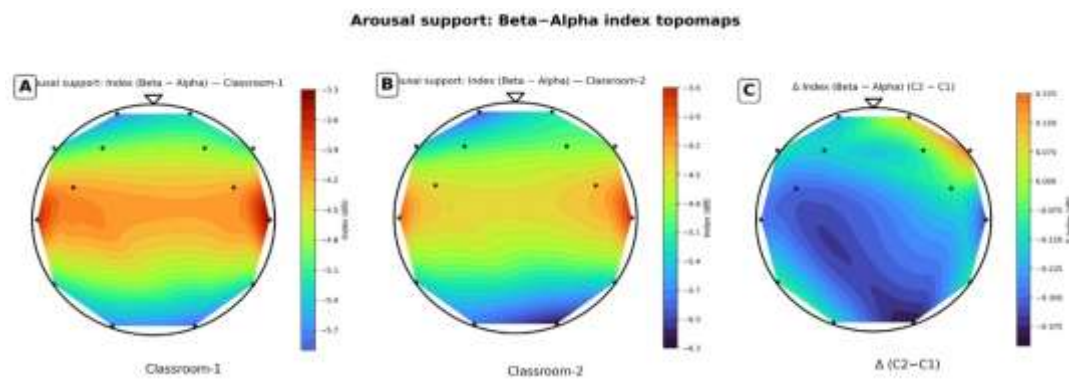


Fig. 12. Arousal support: β - α index topographic maps (C1, C2, Δ).

TABLE VI. AROUSAL SUPPORT: CHANNEL-LEVEL MATCHED STATISTICS FOR THE B-A INDEX TOPOGRAPHIC MAP ($\Delta=C2-C1$).

Channel	n	Δ (C2-C1) mean	t	p	d_z	q_FDR	Sig (q<.05)
O2	24	-0.402	-2.251	0.034	-0.459	0.364	FALSE
FC5	24	-0.299	-2.041	0.052	-0.416	0.364	FALSE
T8	24	-0.359	-1.771	0.089	-0.361	0.415	FALSE
T7	24	-0.292	-1.331	0.196	-0.271	0.649	FALSE
F8	24	0.222	1.205	0.240	0.246	0.649	FALSE
AF3	24	-0.209	-1.110	0.278	-0.226	0.649	FALSE
F3	24	-0.169	-0.833	0.413	-0.170	0.730	FALSE
P8	24	-0.095	-0.707	0.486	-0.144	0.730	FALSE
O1	24	-0.140	-0.651	0.521	-0.132	0.730	FALSE
FC6	24	-0.093	-0.574	0.571	-0.117	0.730	FALSE
F7	24	-0.113	-0.545	0.590	-0.111	0.730	FALSE
P7	24	-0.098	-0.444	0.660	-0.090	0.730	FALSE
AF4	24	0.098	0.420	0.678	0.085	0.730	FALSE
F4	24	-0.069	-0.317	0.753	-0.064	0.753	FALSE

Arousal support β/α ratio

β/α ratio topographic maps showed a similar distribution under both conditions (Fig. 13A-B). The difference map ($\Delta=C2-C1$) remains at very small amplitudes (approximately

± 0.06 on the scale) and does not indicate a strong topographic change (Fig. 13C). Channel statistics also support this finding; no significant channels were found after multiple comparisons (Fig. 13; Table 7).

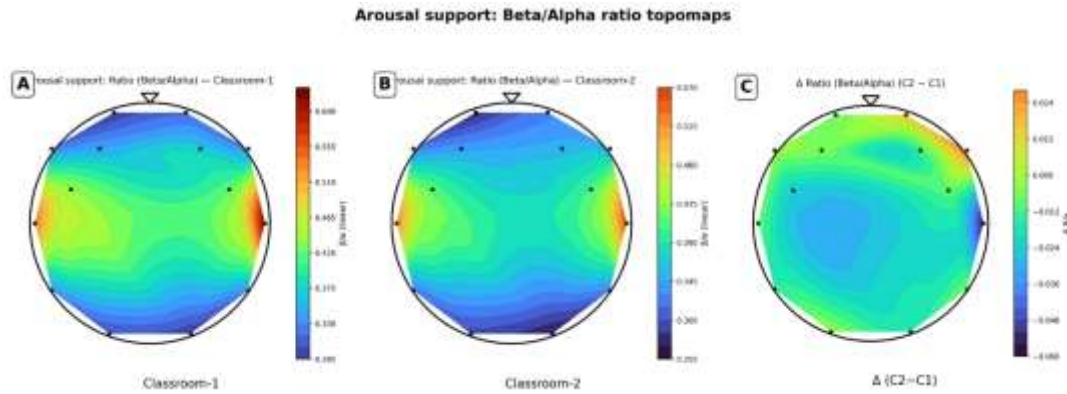


Fig. 13. Arousal support: β/α ratio topographic maps (C1, C2, Δ).

TABLE VII. AROUSAL SUPPORT: CHANNEL-LEVEL MATCHED STATISTICS FOR THE B/A RATIO TOPOGRAPHIC MAP ($\Delta=C2-C1$).

Channel	n	Δ (C2-C1) mean	t	p	d_z	q_FDR	Sig (q<.05)
T8	24	-0.056	-1.591	0.125	-0.324	0.521	FALSE
O2	24	-0.018	-1.545	0.135	-0.315	0.521	FALSE
F8	24	0.029	1.501	0.146	0.306	0.521	FALSE
FC5	24	-0.023	-1.492	0.149	-0.304	0.521	FALSE
P7	24	-0.015	-1.091	0.286	-0.222	0.786	FALSE
AF4	24	0.015	0.814	0.423	0.166	0.786	FALSE
FC6	24	-0.009	-0.695	0.493	-0.141	0.786	FALSE
F4	24	-0.013	-0.686	0.499	-0.14	0.786	FALSE
AF3	24	-0.008	-0.676	0.505	-0.137	0.786	FALSE
P8	24	-0.005	-0.396	0.695	-0.08	0.942	FALSE
F3	24	-0.005	-0.255	0.801	-0.052	0.942	FALSE
T7	24	-0.006	-0.162	0.872	-0.033	0.942	FALSE
O1	24	-0.001	-0.158	0.875	-0.032	0.942	FALSE
F7	24	0	-0.001	0.998	0	0.998	FALSE

In summary, topographic map analyses indicate that asymmetry changes are significant on the valence side in the selected lateral frontal pair (F8-F7), while on the arousal side ($\beta-\alpha$ and β/α), there is no significant increase at the group level. This pattern is consistent with the observation of increased valence in VA results while arousal remains stable [20].

F. Emotiv Performance Metrics

EEG-based VA findings were supplemented by scaled scores (Attention, Engagement, Excitement, Long-Term Excitement, Focus, Interest, Relaxation, Stress) obtained from Emotiv's "Performance Metrics" stream, which were compared between Classroom-1 and Classroom-2 conditions on a matched basis. Participant-based changes and group summaries are presented in Figs. 14-16, while the directions of increase/decrease and statistical tests are presented in Table 8.

Overall, small-amplitude changes were observed between conditions in most metrics; after multiple comparison correction (BH-FDR), no significant differences at the $q < 0.05$

level were detected in any metric. However, in the unadjusted analysis, a downward trend is evident in the Engagement metric in Classroom-2: $M = 0.615$ ($SD = 0.102$) in Classroom-1, $M = 0.590$ ($SD = 0.106$) in Classroom-2; mean difference $\Delta(C2-C1) = -0.025$, 95% CI $[-0.044, -0.007]$. The paired t-test $t(23) = -2.81$, $p = 0.010$ and the Wilcoxon test $p = 0.012$ indicate a medium effect size ($d_z = -0.57$); however, after BH-FDR adjustment, this finding fell below the threshold for statistical significance ($q = 0.080$) (Fig. 16; Table 8).

Similarly, a decreasing pattern was observed in the Long-Term Excitement metric in Classroom-2 ($\Delta = -0.037$, 95% CI $[-0.078, 0.005]$); Although the Wilcoxon test approached significance at the uncorrected level ($p = 0.042$), the paired t-test was not significant ($p = 0.079$) and the result was not preserved after multiple comparisons ($q = 0.284$). Average changes in other metrics (Attention, Excitement, Focus, Interest, Relaxation, Stress) remained small and were not statistically significant (all $q > 0.05$) (Figs. 14-16; Table 8).

Direction analysis (descriptive sign test/increasing-decreasing count) suggested that the change in Engagement tended to be in the decreasing direction (decrease in 18 of 24 participants; $p = 0.023$). The increase-decrease distributions in other metrics were more balanced and did not indicate a systematic trend (Table 8). Participant-based line graphs (Fig. 14) also reveal inter-individual heterogeneity; interpretations are kept cautious due to the secondary/exploratory nature and the algorithmic (proprietary) derivation of these metrics.

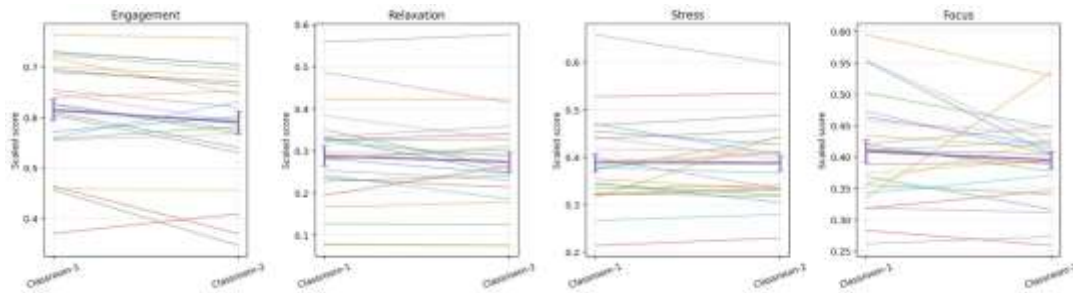


Fig. 14. Participant-based paired changes in selected metrics (Classroom-1→Classroom-2).

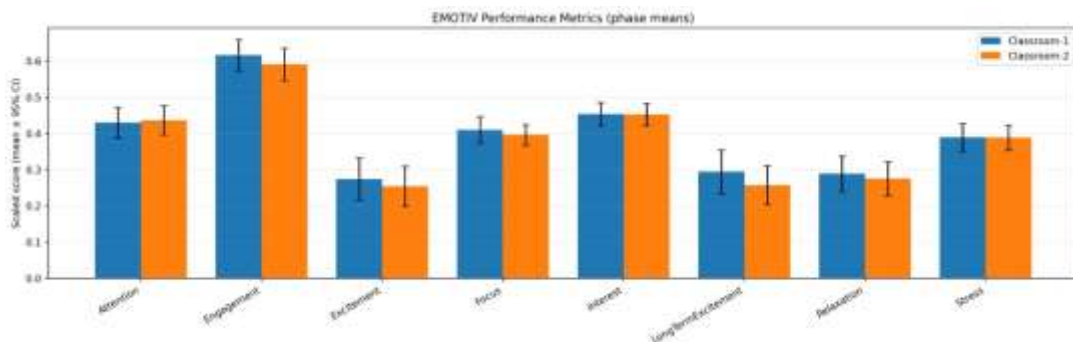


Fig. 15. Condition means (mean $\pm$ 95% CI).

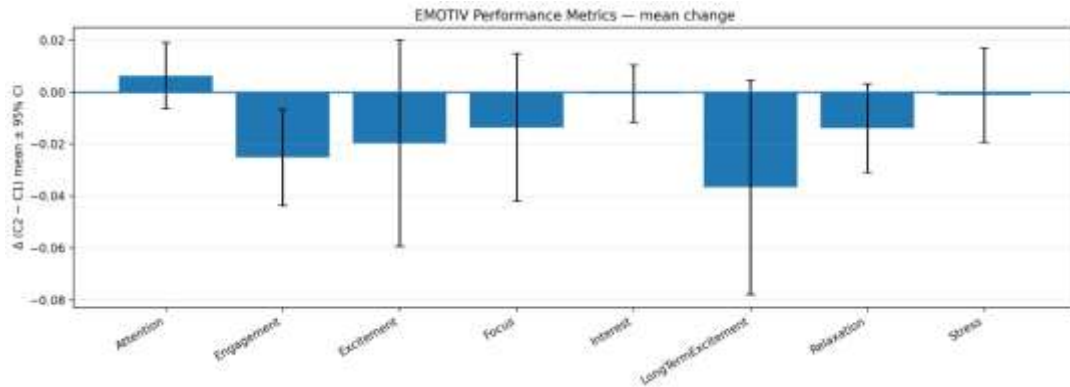


Fig. 16. Average change $\Delta(C2-C1)$ (mean $\pm$ 95% CI).

TABLE VIII. CHANGE DIRECTION (INCREASING/DECREASING), PAIRED TESTS, AND BH-FDR CORRECTED RESULTS FOR EMOTIV PERFORMANCE METRICS.

Metric	Mean Δ (C2-C1)	95% CI Δ	Increase ($\Delta > 0$)	Decrease ($\Delta < 0$)	Sign test p	Paired t-test p	Wilcoxon p	BH-FDR q	d _z
Engagement	-0.025	[-0.044, -0.007]	6	18	0.023	0.010	0.012	0.080	-0.57
Long-Term Excitement	-0.037	[-0.078, 0.005]	9	15	0.307	0.079	0.042	0.284	-0.37
Relaxation	-0.014	[-0.031, 0.003]	10	14	0.541	0.106	0.178	0.284	-0.34
Attention	0.006	[-0.006, 0.019]	13	11	0.839	0.316	0.277	0.445	0.21
Excitement	-0.020	[-0.059, 0.020]	9	15	0.307	0.317	0.439	0.445	-0.21
Focus	-0.014	[-0.042, 0.015]	9	15	0.307	0.334	0.169	0.445	-0.20
Stress	-0.001	[-0.019, 0.017]	12	12	1.000	0.899	0.812	0.905	-0.03
Interest	-0.001	[-0.012, 0.010]	13	11	0.839	0.905	0.900	0.905	-0.02

G. Integrated interpretation and implications for classroom design

Taken together, the findings reveal an instructive divergence between stated preference and task-evoked affect. While students reported strong subjective preference for Classroom-1-driven mainly by lighting and perceived spaciousness-EEG-based VA results indicated higher valence in Classroom-2 during cognitively demanding, exam-like tasks, with arousal remaining stable. This suggests that preference judgments may reflect broad contextual appraisals (comfort expectations, general liking), whereas EEG-derived VA indices may be more sensitive to momentary affective tone under cognitive load. Because the two classrooms differ as integrated spatial bundles (rather than a single manipulated variable), results are best interpreted at the configuration level: educational spaces should be evaluated as “packages” of interacting properties rather than through a single parameter.

H. Limitations and future research

The ecological validity of comparing real classroom environments comes with limited single-factor attribution. Several environmental contributors (e.g., thermal conditions, acoustics, glare, indoor air quality) were not instrumented during sessions. The sample represents a single school context and age group, and behavioral performance outcomes were not positioned as primary endpoints. Future work should include systematic environmental monitoring, controlled feature manipulations to disentangle contributors, broader replication across contexts and tasks, and integrated behavioral outcomes to strengthen causal design inferences.

IV. CONCLUSION

This study compared two classroom environments as integrated spatial configurations and examined students' stated preferences alongside EEG-derived affective indicators during short, cognitively demanding tasks. Survey findings showed a strong preference for Classroom-1, with lighting and perceived spaciousness reported as dominant reasons. In contrast, valence–arousal analysis mapped onto Russell's affect space indicated a more positive valence profile in Classroom-2, while arousal remained broadly stable across conditions. Transition analyses further suggested that this effect reflects a subtle valence shift rather than categorical changes in affective state.

Feature-level summaries based on conventional frontal band-power indices did not yield robust group differences, whereas spatially sensitive analyses highlighted a focal lateral frontal alpha asymmetry effect, pointing to localized neurophysiological sensitivity to classroom configuration. Overall, the results emphasize that subjective preference and task-evoked affect can diverge and support triangulated evaluation of educational environments, combining subjective appraisal with objective, task-based measures, rather than relying on a single indicator or isolated design parameter.

ACKNOWLEDGEMENT

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

REFERENCES

- [1] T. Gunter and J. Shao, "Synthesizing the Effect of Building Condition Quality on Academic Performance," *Educ. Finance Policy*, vol. 11, no. 1, pp. 97–123, Jan. 2016, doi: 10.1162/EDFP_a_00181.
- [2] G. I. Earthman, "School Facility Conditions and Student Academic Achievement," 2002.
- [3] S. I. Higgins, E. Hall, K. Wall, P. Woolner, and C. McCaughey, "The Impact of School Environments : A literature review Produced for the Design Council," 2005.
- [4] C. Cajochen, "Alerting effects of light," *Sleep Med. Rev.*, vol. 11, no. 6, pp. 453–464, Dec. 2007, doi: 10.1016/j.smrv.2007.07.009.
- [5] S. L. Chellappa, R. Steiner, P. Blattner, P. Oelhafen, T. Götz, and C. Cajochen, "Non-Visual Effects of Light on Melatonin, Alertness and Cognitive Performance: Can Blue-Enriched Light Keep Us Alert?," *PLoS One*, vol. 6, no. 1, p. e16429, Jan. 2011, doi: 10.1371/journal.pone.0016429.
- [6] P. Barrett, F. Davies, Y. Zhang, and L. Barrett, "The impact of classroom design on pupils' learning: Final results of a holistic, multi-level analysis," *Build. Environ.*, vol. 89, pp. 118–133, Jul. 2015, doi: 10.1016/j.buildenv.2015.02.013.
- [7] I. Bower, R. Tucker, and P. G. Enticott, "Impact of built environment design on emotion measured via neurophysiological correlates and subjective indicators: A

- systematic review," *J. Environ. Psychol.*, vol. 66, p. 101344, Dec. 2019, doi: 10.1016/j.jenvp.2019.101344.
- [8] M. Llorens-Gómez, J. L. Higuera-Trujillo, C. S. Omarrementeria, and C. Llinares, "The impact of the design of learning spaces on attention and memory from a neuroarchitectural approach: A systematic review," *Frontiers of Architectural Research*, vol. 11, no. 3, pp. 542–560, Jun. 2022, doi: 10.1016/j.foar.2021.12.002.
- [9] J. Meyers-Levy and R. Zhu, "The Influence of Ceiling Height: The Effect of Priming on the Type of Processing That People Use," *Journal of Consumer Research*, vol. 34, no. 2, pp. 174–186, Aug. 2007, doi: 10.1086/519146.
- [10] O. Vartanian et al., "Architectural design and the brain: Effects of ceiling height and perceived enclosure on beauty judgments and approach-avoidance decisions," *J. Environ. Psychol.*, vol. 41, pp. 10–18, Mar. 2015, doi: 10.1016/j.jenvp.2014.11.006.
- [11] G. Franz, M. von der Heyde, and H. H. Bühlhoff, "An empirical approach to the experience of architectural space in virtual reality—exploring relations between features and affective appraisals of rectangular indoor spaces," *Autom. Constr.*, vol. 14, no. 2, pp. 165–172, Mar. 2005, doi: 10.1016/j.autcon.2004.07.009.
- [12] İ. Erkan, "Examining wayfinding behaviours in architectural spaces using brain imaging with electroencephalography (EEG)," *Archit. Sci. Rev.*, vol. 61, no. 6, pp. 410–428, Nov. 2018, doi: 10.1080/00038628.2018.1523129.
- [13] H. Laufs et al., "EEG-correlated fMRI of human alpha activity," *Neuroimage*, vol. 19, no. 4, pp. 1463–1476, Aug. 2003, doi: 10.1016/S1053-8119(03)00286-6.
- [14] D. Metzen, E. Genç, S. Getzmann, M. F. Larra, E. Wascher, and S. Ocklenburg, "Frontal and parietal EEG alpha asymmetry: a large-scale investigation of short-term reliability on distinct EEG systems," *Brain Struct. Funct.*, vol. 227, no. 2, pp. 725–740, Mar. 2022, doi: 10.1007/s00429-021-02399-1.
- [15] J. J. B. Allen, P. M. Keune, M. Schönenberg, and R. Nusslock, "Frontal EEG alpha asymmetry and emotion: From neural underpinnings and methodological considerations to psychopathology and social cognition," *Psychophysiology*, vol. 55, no. 1, Jan. 2018, doi: 10.1111/psyp.13028.
- [16] A. T. Pope, E. H. Bogart, and D. S. Bartolome, "Biocybernetic system evaluates indices of operator engagement in automated task," *Biol. Psychol.*, vol. 40, no. 1–2, pp. 187–195, May 1995, doi: 10.1016/0301-0511(95)05116-3.
- [17] F. G. Freeman, P. J. Mikulka, L. J. Prinzl, and M. W. Scerbo, "Evaluation of an adaptive automation system using three EEG indices with a visual tracking task," *Biol. Psychol.*, vol. 50, no. 1, pp. 61–76, May 1999, doi: 10.1016/S0301-0511(99)00002-2.
- [18] R. Ramirez and Z. Vamvakousis, "Detecting Emotion from EEG Signals Using the Emotive Epoc Device," 2012, pp. 175–184. doi: 10.1007/978-3-642-35139-6_17.
- [19] I. Marcantoni et al., "Ratio Indexes Based on Spectral Electroencephalographic Brainwaves for Assessment of Mental Involvement: A Systematic Review," *Sensors*, vol. 23, no. 13, p. 5968, Jun. 2023, doi: 10.3390/s23135968.
- [20] J. A. Russell, "A circumplex model of affect," *J. Pers. Soc. Psychol.*, vol. 39, no. 6, pp. 1161–1178, Dec. 1980, doi: 10.1037/h0077714.
- [21] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.
- [22] A. Mollahosseini, B. Hasani, and M. H. Mahoor, "AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild," *IEEE Trans. Affect. Comput.*, vol. 10, no. 1, pp. 18–31, Jan. 2019, doi: 10.1109/TAFFC.2017.2740923.
- [23] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, no. 3–4, pp. 591–611, Dec. 1965, doi: 10.1093/biomet/52.3-4.591.

- [24] F. Wilcoxon, "Individual Comparisons by Ranking Methods," *Biometrics Bulletin*, vol. 1, no. 6, p. 80, Dec. 1945, doi: 10.2307/3001968.
- [25] D. Lakens, "Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs," *Front. Psychol.*, vol. 4, 2013, doi: 10.3389/fpsyg.2013.00863.
- [26] H. Hotelling, "The Generalization of Student's Ratio," *The Annals of Mathematical Statistics*, vol. 2, no. 3, pp. 360–378, Aug. 1931, doi: 10.1214/aoms/1177732979.
- [27] Y. Benjamini and Y. Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *J. R. Stat. Soc. Series B Stat. Methodol.*, vol. 57, no. 1, pp. 289–300, Jan. 1995, doi: 10.1111/j.2517-6161.1995.tb02031.x.
- [28] A. H. Bowker, "A Test for Symmetry in Contingency Tables," *J. Am. Stat. Assoc.*, vol. 43, no. 244, pp. 572–574, Dec. 1948, doi: 10.1080/01621459.1948.10483284.
- [29] J. Cohen, "A Coefficient of Agreement for Nominal Scales," *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 37–46, Apr. 1960, doi: 10.1177/001316446002000104.
- [30] A. E. Maxwell, "Comparing the Classification of Subjects by Two Independent Judges," *British Journal of Psychiatry*, vol. 116, no. 535, pp. 651–655, Jun. 1970, doi: 10.1192/bjp.116.535.651.
- [31] Q. McNemar, "Note on the Sampling Error of the Difference Between Correlated Proportions or Percentages," *Psychometrika*, vol. 12, no. 2, pp. 153–157, Jun. 1947, doi: 10.1007/BF02295996.
- [32] A. Stuart, "A Test for Homogeneity of the Marginal Distributions in a Two-Way Classification," *Biometrika*, vol. 42, no. 3/4, p. 412, Dec. 1955, doi: 10.2307/2333387.
- [33] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*. Chapman and Hall/CRC, 1994. doi: 10.1201/9780429246593.
- [34] R. J. Davidson, "Anterior cerebral asymmetry and the nature of emotion," *Brain Cogn.*, vol. 20, no. 1, pp. 125–151, Sep. 1992, doi: 10.1016/0278-2626(92)90065-T.
- [35] J. A. Coan and J. J. B. Allen, "Frontal EEG asymmetry as a moderator and mediator of emotion," *Biol. Psychol.*, vol. 67, no. 1–2, pp. 7–50, Oct. 2004, doi: 10.1016/j.biopsycho.2004.03.002.
- [36] W. Klimesch, "EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis," *Brain Res. Rev.*, vol. 29, no. 2–3, pp. 169–195, Apr. 1999, doi: 10.1016/S0165-0173(98)00056-3.