Submitted to **Journal Title**
Template provided by EDP Sciences

RESEARCH ARTICLE

Effects of response scale conversion and questionnaire configuration on soundscape assessment based on ISO/TS 12913-3

Lucas Rafael Ferreira[1⋆,⋆⋆], Leonardo Marques Monteiro[2], Nara Gabriela de Mesquita Peixoto[3,⋆⋆], and Ranny Loureiro Xavier Nascimento Michalski[4,⋆⋆]

[1] University of São Paulo, lucasrafael2209@usp.br
[2] University of São Paulo, leo4mm@usp.br
[3] University of São Paulo, nara.peixoto@usp.br
[4] University of São Paulo, rannym@usp.br

Abstract – Soundscape evaluation commonly relies on Perceived Affective Quality (PAQ) questionnaires to translate subjective auditory experience into structured perceptual data. However, while ISO/TS 12913-2 provides multiple standardized data collection options, the comparative effects of different response scale formats and PAQ questionnaire configurations on statistical outcomes and circumplex representations have not been systematically quantified in applied soundscape studies. This study investigates how these methodological decisions affect both statistical outcomes and the soundscape circumplex representation defined in ISO/TS 12913-3. A field experiment was conducted across five acoustically heterogeneous university environments in São Paulo, Brazil, involving 246 completed questionnaires from four independent groups. Four questionnaire versions were compared by combining two response scale formats (five-point ordinal versus continuous metric 0–100) and two PAQ configurations (four bipolar adjective pairs versus eight single attributes). Ordinal responses were converted into metric values using Snell's latent-scale transformation and a fixed-value mapping approach. The resulting datasets were analyzed using Mann–Whitney U tests, descriptive statistics, and kernel-density circumplex graphs. Snell's conversion led to attribute-specific redistributions, notably affecting medians and kurtosis, whereas fixed-value mapping closely aligned with native metric responses. Questionnaire configuration exerted a systematic influence on perceptual outcomes: bipolar PAQs produced compact distributions concentrated in the Calm quadrant, while the eight-attribute configuration captured broader perceptual variability extending toward the Vibrant and Chaotic regions. The attributes Uneventful and Monotonous exhibited the highest distributional instability across configurations. The convergence between inferential statistics and circumplex representations reinforces the validity of the results. Overall, the findings demonstrate that response scale conversion and PAQ configuration critically shape ISO-based soundscape interpretations, providing empirically grounded methodological guidance for harmonizing PAQ applications in perceptual-acoustic research.

Keywords. ISO 12913, Perceived Affective Quality, Psychometric analysis, Response scale format, Soundscape

1. Introduction

Rapid urbanization has increased the need to evaluate environmental quality and human well-being in cities [1]. Consequently, research in urban planning, architecture, and environmental engineering seeks methodologies that integrate physical attributes with subjective experience [2]. Within this context, the soundscape approach has emerged as a standardized framework for assessing acoustic environments, moving beyond traditional analyses based solely on sound pressure level (SPL). Recent work on urban acoustic environments likewise indicates that sound pressure level indices alone may overlook relevant temporal, spectral, and source-related structures, motivating multimetric characterizations that complement SPL with additional acoustic indicators [3].

By integrating perceptual and contextual dimensions, the soundscape approach provides a comprehensive understanding of human interaction with acoustic environments [4–7]. Environmental sound is recognized as a key element of the multisensory experience of a place, influencing how people affectively and cognitively evaluate urban spaces [8]. Consequently, this perspective has led to

⋆ Corresponding author: lucasrafael2209@usp.br
⋆⋆ These authors contributed equally to this work.

This provisional PDF is the accepted version. The article should be cited as: Acta Acustica, https://doi.org/10.1051/aacus/2026046

the development of standardized perceptual assessment tools that translate subjective acoustic experience into structured data.

The ISO 12913 series [9–11] formalizes the soundscape paradigm, emphasizing that soundscape is the acoustic environment as perceived or experienced and/or understood by a person or people, in context, and that context comprises the interrelationships between person, activity, and place, in space and time, which may influence auditory sensation, interpretation, and responses to the acoustic environment [4, 12–14]. The ISO 12913 series provides an internationally accepted framework for soundscape research: Part 1 [9] defines the concept and model, Part 2 [10] details in situ data collection and reporting, and Part 3 [11] guides data analysis. In practical implementations of ISO-guided soundscape assessments, researchers must still make specific design choices regarding questionnaire structure and response formats (e.g., bipolar adjective pairs versus single-attribute items, and ordinal versus metric response options). Such choices can systematically influence how perceptual data are summarized and interpreted in subsequent analyses. However, the extent to which these design decisions affect distributional outcomes and circumplex-based representations has not been consistently quantified in applied soundscape studies. This study therefore addresses these methodological issues, particularly regarding response-scale conversion and questionnaire configuration.

Among the most widely adopted instruments for perceptual soundscape assessment is the Perceived Affective Quality (PAQ) questionnaire, composed of bipolar adjective pairs (e.g., *Calm – Chaotic*, *Monotonous – Vibrant*) [10, 15, 16]. These instruments are essential in soundscape research because they provide standardized methods for collecting perceptual data, which may inform evidence-based planning and policymaking in environmental acoustics and urban design [17–20].

Despite the availability of standardized PAQ formats in ISO/TS 12913-2, the comparative effects of these methodological options have not yet been systematically quantified. Specifically, how do different response scale formats and questionnaire configurations influence statistical outcomes, distributional properties, and circumplex-based representations? Understanding these effects is essential for researchers to make informed methodological choices aligned with their analytical objectives. In particular, the distinction between metric and ordinal scales, especially when applied to bipolar semantic configurations, has received limited attention in the soundscape literature [21–24]. However, the choice of scale format may substantially affect not only the sensitivity and resolution of responses, but also the cognitive effort required from participants, the clarity of item interpretation, and the consistency of affective judgments [15, 24–27].

The existing literature on noise annoyance [28, 29], particularly from the International Commission on Biological Effects of Noise (ICBEN), describes two widely used response formats: a five-point ordinal scale with verbal labels, and a 0–10 point numeric rating scale adapted for different languages. Brink *et al.* [30] converted both the 5-point ordinal and 11-point numeric rating scales to a 0–100 metric, mapping verbal categories to 6, 25, 52, 79, and 95. This conversion underestimated annoyance at the lower end (0 vs. 6) and overestimated it at the upper end (100 vs. 95). Nevertheless, the authors endorsed the ICBEN conversion, as no alternative method offered clear advantages at the extremes. Similarly, Zhang *et al.* [31] compared a 0–100 linear scale with a 0–10 category scale containing five verbal anchors in laboratory listening tests and found that perceived annoyance results were largely independent of scale type.

From a statistical standpoint, metric scales (interval or ratio) yield numerical values with consistent intervals, making them compatible with a broad range of statistical analyses, including parametric techniques when distributional assumptions are sufficiently met. Such scales enable powerful and detailed modeling across many contexts. Ordinal scales, in contrast, represent ordered categories without defined distances between them. They can simplify the response process and reduce interpretative ambiguity, but they restrict the analyst to ordinal-specific or nonparametric methods, limiting the scope and complexity of statistical models that can be applied [32].

Moreover, when semantic polarity is introduced through bipolar adjective pairs, the interaction between scale format and item configuration becomes even more critical. It remains unclear whether certain scale types facilitate or hinder expressing perceptual contrasts, particularly when respondents are asked to make nuanced affective distinctions [33, 34].

While ISO/TS 12913-2 [10] provides standardized options for scale response formats and questionnaire configurations, the comparative effects of selecting among these options on downstream analytical outcomes have not yet been systematically quantified in applied soundscape studies, which limits the empirical basis for recommending specific format–configuration combinations in different research contexts. This concern is consistent with broader observations in soundscape research, where data-collection tools and protocols vary substantially across studies, particularly with respect to the use of questionnaires and semantic scales in different empirical designs [16]. Moreover, survey methodology has repeatedly shown that response scale format and question construction can alter response behavior and, consequently, statistical results and conclusions, underscoring the importance of empirically testing these design decisions rather than assuming equivalence [24]. Addressing this gap, the present study experimentally compares two methods for converting ordinal responses into a continuous metric scale (Snell's conversion [35] and fixed-value mapping [30]) as well as two PAQ questionnaire configurations (four bipolar PAQs versus eight single-attribute PAQs [24, 25]). Their influence on both statistical results and spatial representations within the ISO/TS 12913-3 [11] soundscape evaluation framework is investigated.

Building on this comparative and experimental approach, the study aims to evaluate the effectiveness, sensitivity, and interpretative clarity of different PAQ response formats in soundscape questionnaires. By integrating statistical metrics with circumplex graph analyses, this research provides methodological guidance for optimizing the design of perceptual instruments in soundscape studies. Circumplex representations were constructed following ISO/TS 12913-3 coordinate transformations [11], applied at the individual response level as recommended by Mitchell *et al.* [36], with visualizations based on mean values of the resulting continuous ISOPleasant and ISOEventful coordinates. Ultimately, it contributes to the development of robust and interpretable data collection tools that deepen the understanding of subjective dimensions of acoustic environments and support evidence-based practices in environmental acoustics and urban perception research.

2. Research Questions and Hypotheses

This study addresses two primary research questions:

RQ1: How do different methods for converting ordinal responses to metric scales (Snell's latent-scale transformation versus fixed-value mapping) affect the statistical properties and circumplex representations of soundscape data?

RQ2: How does questionnaire configuration (four bipolar adjective pairs versus eight single-attribute PAQs) influence perceptual outcomes and distributional characteristics in soundscape assessment?

Based on prior research on scale format effects and the theoretical foundations of the ISO/TS 12913-3 circumplex model, this study advances the following hypotheses:

H1 (scale conversion): Snell's latent-scale transformation will produce attribute-specific redistributions in central tendency and distributional shape, whereas fixed-value mapping will yield outcomes more closely aligned with native metric scales.

H2 (questionnaire configuration): Bipolar PAQ configurations will generate more compact distributions concentrated in specific circumplex quadrants, whereas eight single-attribute configurations will capture broader perceptual variability across the circumplex space.

H3 (cross-analytical convergence): Statistically significant differences identified through non-parametric inferential tests will correspond to observable and directionally consistent shifts in KDE-based circumplex representations, including centroid relocation, contour deformation, and density redistribution across quadrants, providing convergent evidence from independent analytical planes.

3. Material and Methods

The methodological design combined in situ perceptual data collection in selected urban areas with psychometric analysis[1], enabling an investigation of how response scale format (metric versus ordinal) and semantic configurations influence participants' affective judgments. The adopted analytical workflow followed a sequence of procedures designed to ensure the reliability, comparability, and interpretability of the data.

This study was approved by the Research Ethics Committee (CEP) of the University of São Paulo (CAEE 80357224.1.0000.5390 and 63954322.4.0000.5390).

3.1. Study context and site selection

To explore how different response scale formats influence the perceptual evaluation of soundscapes, this study was conducted across five indoor and outdoor environments on the University of São Paulo (USP) campus. Located in the western zone of São Paulo city, Brazil, the campus includes extensive green infrastructure and a mix of natural and built elements that create varied acoustic conditions. Spatial distribution and site characteristics (P1–P5) are shown in Figure 1.

Outdoor sites (P1 and P4) are open green areas with ground-level vegetation, scattered trees, and nearby buildings. Both are adjacent to roads with regular vehicular traffic, introducing dynamic acoustic layers despite predominantly vegetated surroundings. Indoor sites included a large entrance hall (P2), a semi-open transitional area connecting indoor and outdoor spaces (P3), and a classroom (P5). These locations differ in spatial configuration, materiality, and enclosure, all factors that can shape the acoustic experience and users' perceptions.

Site selection followed the principle of environmental heterogeneity. Including spaces with distinct physical, visual, and acoustic characteristics created diverse perceptual contexts for testing response behavior under different real-world urban conditions. This diversity ensured that comparisons between metric and ordinal response formats captured the representative perceptual variability of real soundscapes.

3.2. Questionnaire design and semantic structure

The experimental study followed method A from ISO/TS 12913-2 [10] for in situ soundscape evaluation. All questionnaire versions were based on the ISO soundscape circumplex model, which operationalizes perceptual assessment through semantic differential items. The PAQ attributes (*Pleasant* – *Annoying*, *Eventful* – *Uneventful*, *Vibrant* – *Monotonous*, and *Calm* – *Chaotic*) were used either as four bipolar adjective pairs or as eight single attributes. The response scale formats used across questionnaire configuration are summarized in Figure 2.

Four questionnaire versions were developed by combining two response scale formats (1) and two PAQ questionnaire configurations (2):

1. Five-category ordinal scale Likert-type and continuous metric scale (0–100).
2. Eight single-attribute PAQs and four bipolar adjective pairs.

[1] In the context of this study, psychometric analysis refers to the set of statistical procedures and tests used to assess the reliability, consistency, and structural properties of the questionnaire formats applied to measure perceived affective qualities (PAQ) in urban soundscapes.

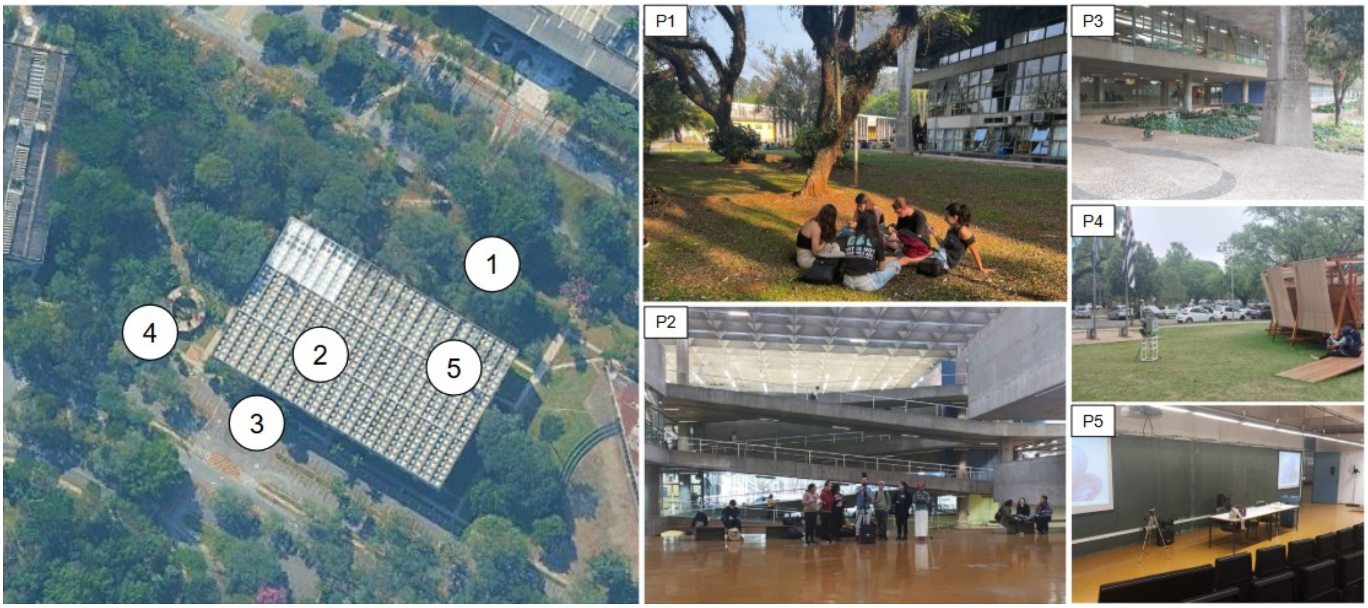

Figure 1. Spatial distribution and visual characterization of the five selected locations.

The field experiment was conducted on September 23 and October 14, 2024, corresponding to two application sessions in which all questionnaire versions were administered under comparable environmental conditions. Four groups were involved, as detailed below. On each application day, 126 students participated and were divided into two groups, each assigned to one questionnaire version. Six responses from Groups 3 and 4 were subsequently excluded due to incomplete or inconsistent data.

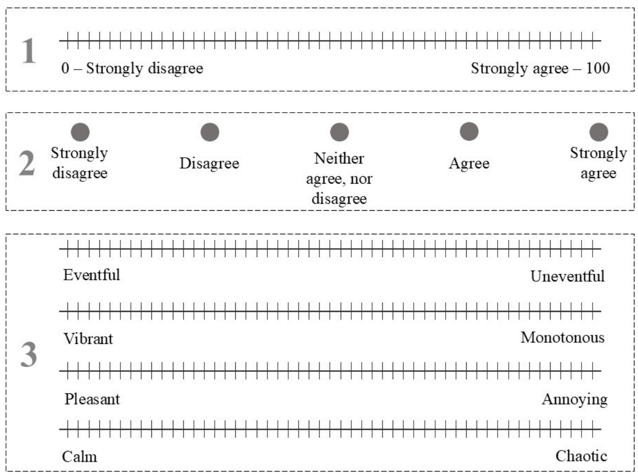

Figure 2. Response scales used across questionnaire configurations: (1) eight single-attributes PAQ metric scale; (2) eight single-attributes PAQ ordinal scale; and (3) four bipolar adjective pairs (PAQ) metric scale.

- Group 1 (G1): $n = 71$, eight single attributes, ordinal response scale. For analytical purposes, this group was later subdivided into `G1_S`, with responses converted using Snell's method, and `G1_i`, with responses converted using fixed-value mapping (see Section 3.4.1).
- Group 2 (G2): $n = 55$, eight single-attribute PAQs, metric response scale.
- Group 3 (G3): $n = 61$, eight single-attribute PAQs, metric response scale.
- Group 4 (G4): $n = 59$, four bipolar adjective pairs, metric response scale.

Data were collected in situ using participants' smartphones through KoboToolbox online platform [37]. All participants received standardized written and verbal instructions to ensure a consistent understanding of the response procedure and provided written informed consent in accordance with the approved ethical procedures.

Finally, to evaluate methodological effects, the dataset was organized into predefined comparison sets (detailed in section 3.4.2) that examined how response scale conversion and PAQ questionnaire configuration influenced perceptual judgments, as summarized in Table 1.

Taken together, these comparisons offer a comprehensive framework for assessing the sensitivity of PAQ instruments to changes in questionnaire design, contributing to a deeper understanding of how perceptual judgments are shaped by instrument structure in soundscape research.

3.3. Data collection and sample characteristics

Data collection was conducted across four distinct application groups (G1 to G4), each corresponding to a distinct questionnaire version as defined in Section 3.2. This design ensured experimental control and enabled direct systematic comparisons between response scale formats and PAQ questionnaire configurations.

Participants were undergraduate students enrolled in the Environmental Comfort course at the Faculty of

Table 1. Comparisons made between the different groups. Comparisons A1 and A2 were conducted at Location P1 (open green area); B1 at P2 (entrance hall); B2 at P3 (transitional area); B3 at P4 (open green area); and B4 at P5 (classroom)

.

Comparison	Groups involved	Scale type	PAQ questionnaire transformation	Location	Objective
A1	G1_S (Snell) and G2	Ordinal scale (verbal labels, Snell transformation) vs. metric scale (numeric values, 0–100)	Eight PAQs	P1	Evaluate the influence of response scale type on perceptual evaluations
A2	G1_i (fixed-value) and G2	Ordinal scale (verbal labels, fixed-value mapping) vs. metric scale (numeric values, 0–100)			
B1	G3 and G4	Metric scale (numeric values, 0–100)	Eight PAQs × Four bipolar PAQs	P2	Investigate the impact of the semantic structure of the PAQs on response interpretation
B2				P3	
B3				P4	
B4				P5	

Architecture, Urbanism, and Design at the University of São Paulo (FAU-USP). This context provided access to individuals with foundational knowledge of environmental quality, supporting informed and contextually grounded soundscape evaluations.

Participation was voluntary (convenience sampling) and followed inclusion criteria requiring individuals to be at least 18 years old and reporting no hearing impairments. All participants were aged between 18 and 25 years. The sample encompassed individuals with diverse personal and perceptual backgrounds, consistent with the subjective nature of soundscape assessment. Such heterogeneity is methodologically relevant, as individual sensitivities and spatial interpretations inherently shape perceptual judgments within the ISO 12913 framework.

Although group sizes varied, all samples were adequate to support the psychometric and comparative analyses applied, as shown in Table 1. Table 2 summarizes the demographic characteristics of the analyzed groups.

Table 2. Demographic data of the participants.

Group	n	Female (%)	Male (%)
G1	71	56.34	43.66
G2	55	61.82	38.18
G3	61	63.16	36.84
G4	59	75.86	24.14

3.4. Statistical procedures and comparative analysis

All statistical analyses were carried out using R (v. 4.5.0) [38] and JASP (v. 0.19.3.0) [39], while visual representations were generated with *Soundscapy*, a Python toolbox developed by Mitchell *et al.* [36] for perceptual-acoustic soundscape analysis. Following ISO/TS 12913-3 recommendations [11], median values are reported as the primary measure of central tendency for ordinal and ordinal-derived data, given that ordinal scales represent ordered categories without defined intervals. Mean values are additionally reported to provide supplementary distributional context and to facilitate comparison with metric-scale data. Standard deviation, skewness, and kurtosis are also reported to characterize distributional variability, asymmetry, and concentration. This approach ensures that statistical analyses respect the measurement properties of each scale format while enabling comprehensive characterization of response distributions.

3.4.1. Response scale transformation

Given the dual nature of the response formats, ordinal and metric, the first analytical step involved converting ordinal data into continuous values to enable cross-format comparisons under consistent methodological assumptions. Two conversion strategies were applied in parallel.

The first strategy employed Snell's method [35, 40–42]. Because the questionnaire version used for comparison employed a native 0–100 scale, this conversion was selected to enable direct cross-format comparisons within a common numerical range. Accordingly, a monotonic transformation was prioritized to preserve category ordering while avoiding *ad hoc* assumptions of equal spacing between ordinal categories. Snell's procedure estimates latent category thresholds from ordinal data using the inverse logit function[2], and each ordinal response is then assigned a continuous latent score corresponding to the midpoint between adjacent thresholds. The resulting latent scores were then linearly rescaled to the 0–100 interval for numerical alignment with the metric format.

[2] The inverse logit function, also known as the logistic function, is a mathematical function that transforms a real number into a value between 0 and 1.

The second strategy consisted of assigning fixed numerical values to ordinal categories, following established practices in perceptual research [43, 44]. In this approach, response options were mapped as follows: 0 (strongly disagree), 25 (disagree), 50 (neither agree, nor disagree), 75 (agree), and 100 (strongly agree). This conventional mapping provided a pragmatic basis for cross-format comparisons, while maintaining interpretative clarity for respondents and analytical consistency for researchers.

For the four bipolar PAQ attributes *Pleasant – Annoying*, *Eventful – Uneventful*, *Vibrant – Monotonous*, and *Calm – Chaotic*, a complementary conversion procedure was applied to ensure analytical symmetry and comparability across PAQs. Participants rated each pair using a unidirectional 0–100 bipolar scale. The original score was retained for the left-hand term (e.g., *Pleasant*), while its complement to 100 was assigned to the opposite pole (e.g., *Annoying*). Thus, a rating of 20 on the *Pleasant – Annoying* scale was interpreted as 20 for *Pleasant* and 80 for *Annoying*.

This procedure produced eight complementary variables from the four bipolar adjective pairs, enhancing interpretability and enabling direct comparison across PAQ questionnaire configurations and scale formats.

3.4.2. Group comparisons

After the conversion procedures, the four comparisons listed on Table 1 were conducted.

Comparisons A1 and A2 contrasted Group 1, whose ordinal responses were converted using either Snell's method or fixed-value mapping, with Group 2, which employed a native metric scale (0–100) under the same questionnaire configuration. Comparisons B1 through B4 examined Groups 3 and 4, both using metric formats but differing in PAQ configurations (eight single attributes vs. four bipolar adjective pairs).

Because the converted outcomes displayed non-normal distributions, between-group differences were assessed using the Mann–Whitney U (Wilcoxon rank-sum) test [45, 46]. The test is computed from the pooled ranking of all observations and the sum of ranks within each group, with U being a one-to-one transformation of the rank-sum statistic [46]. In this framework, the "ranked means" commonly reported in software outputs correspond to the mean rank, defined as the rank sum divided by the group sample size, and provide a descriptive indication of which group tends to occupy higher positions in the pooled ordering [46]. Importantly, the Mann–Whitney test should be interpreted as assessing whether one group tends to yield systematically higher values than the other, that is, whether the samples are consistent with coming from the same distribution, rather than as a general test of equal medians [46, 47]. A median-based interpretation is only justified under a location-shift scenario in which the two population distributions have similar shapes; otherwise, statistically significant results may also reflect differences in spread or distributional shape [47–49].

Standard assumptions for application in the present design, including independent samples and outcomes measured on an ordinal or continuous scale, were verified before analysis [50].

3.4.3. Kernel-density circumplex analysis

To complement the statistical analyses, two-dimensional perceptual soundscape circumplex graphs were generated for each group, representing the dimensions of *Eventfulness* (E) and *Pleasantness* (P) according to the ISO/TS 12913-3 framework [11]. The graphs were constructed using *Soundscapy* [36] and were based on participants' metric-scale responses. Kernel density estimation employed Scott's rule for automatic bandwidth selection, as implemented in Soundscapy's default configuration, ensuring data-driven smoothing consistent with the sample sizes in each group.

Dimensional values were calculated using the standardized equations (Eq. 1 and Eq. 2) normalized from -1 to $+1$. The equations remain applicable to bipolar scales, as each response inherently represents the perceptual distance between opposite poles.

$$P = (p - a) + \cos 45°.\,(ca - ch) + \cos 45°.\,(v - m) \quad (1)$$

$$E = (e - u) + \cos 45°.\,(ch - ca) + \cos 45°.\,(v - m) \quad (2)$$

where: p is pleasant, a is annoying, ca is calm, ch is chaotic, v is vibrant, m is monotonous, e is eventful, and u is uneventful.

As a final step, the statistical and spatial results were jointly examined to assess how different analytical configurations influence PAQ data interpretation and to identify context-specific recommendations for applying these methods in urban soundscape research.

4. Results

4.1. Statistical analysis

This section presents the statistical outcomes from the comparative analyses evaluating how response scale format (Comparisons A) and PAQ questionnaire configuration (Comparisons B) affect PAQ ratings. Between-group differences are assessed using non-parametric significance tests, adopting the conventional significance level ($\alpha = 0.05$, two-sided), and exact p-values are reported to ensure transparency and support interpretation. To contextualize inferential results, Tables 3 to 8 also report descriptive statistics for each PAQ: mean (M) and median (Mdn) as complementary indicators of central tendency, standard deviation (SD) as a measure of dispersion, skewness (g_1) to quantify distribution asymmetry, and kurtosis (g_2) to characterize distribution concentration and tail behavior. Reporting these descriptors jointly is particularly relevant given the ordinal origins and non-normal distribution patterns associated with the transformed datasets, allowing statistical significance to be interpreted alongside shifts in location, spread, and distributional shape.

4.1.1. Comparisons A

Comparisons A evaluated whether the response scale format, namely an ordinal scale converted to a continuous metric or a native metric scale, significantly affected PAQ ratings.

Comparison A1 (Table 3) contrasted an ordinal scale converted using Snell's method (G1_S) with a native metric (0–100) scale (G2). Under the Snell conversion, the PAQs *Vibrant*, *Pleasant*, and *Chaotic* exhibited the greatest divergence, indicating redistribution of responses along the scale range. For the remaining attributes, no statistically significant differences were observed (all p-values > 0.05), suggesting that, within the limits of the sample size and test sensitivity, ratings were not distinguishable between the two scale formats.

Mann–Whitney U tests revealed statistically significant differences for *Vibrant* (U = 1146.00, p-value = 0.003), *Pleasant* (U = 2379.50, p-value < 0.001), and *Chaotic* ($U = 1127.00$, p-value$=0.002$). Table 3 presents the Mann–Whitney U results and the descriptive statistics (M, Mdn, SD, g_1, and g_2) for all PAQs in Comparison A1, reporting values as G1_S|G2.

Specifically, for *Vibrant*, the Snell-converted format (G1_S) tended to yield lower values than the metric format, as reflected in median values (10.80 vs. 20.75), and a smaller standard deviation ($SD = 17.29$ vs. 23.38). Conversely, *Pleasant* tended toward higher values in G1_S (medians of 91.38 vs. 76.50) with lower dispersion ($SD = 12.89$ vs. 18.73). For *Chaotic*, G2 tended to produce higher ratings, with medians of 17.50 versus 9.46, with comparable dispersion ($SD = 17.43$ vs. 19.75).

Differences in kurtosis were also observed among these significant PAQs, for example, *Pleasant* exhibited a more peaked distribution in G1_S (1.14) than G2 (0.12), suggesting reduced response variability under the Snell conversion. Skewness values also shifted under Snell's conversion, for instance *Pleasant* became more negatively skewed in G1_S than in G2 (–1.40 vs. –0.73), indicating stronger concentration toward higher ratings.

In Comparison A2 (Table 4), which contrasted the fixed-value mapping (0, 25, 50, 75, 100), G1_i, with the native metric scale, G2, no statistically significant differences were identified for any PAQ (all p-values > 0.05). Descriptive measures appeared similar across formats. For instance, *Calm* had the same median in both formats (75.00), and *Pleasant* medians were close (75.00 vs. 76.50).

To quantify the magnitude of differences between response formats, the rank-biserial correlation (r_{rb}) was employed as an effect size measure for the Mann-Whitney U test. It is computed as $r_{rb} = 2U/(n_1 \times n_2) - 1$, where U is the Mann-Whitney statistic and n_1 and n_2 are the sample sizes of the two groups under comparison. The resulting values range from -1 to $+1$, where 0 indicates no difference between distributions and values approaching ± 1 reflect near-complete separation between groups. Following conventional benchmarks, effect sizes were classified as small ($|r_{rb}| \geq 0.1$), medium ($|r_{rb}| \geq 0.3$), and large ($|r_{rb}| \geq 0.5$). The sign of r_{rb} is informative regarding the direction of differences: positive values indicate that n_2 tends to produce higher ratings for a given attribute, while negative values indicate the opposite.

Effect sizes for Comparisons A were generally small to negligible for both conversion methods, with the exception of *Vibrant* ($r_{rb} = 0.413$) and *Chaotic* ($r_{rb} = 0.423$) in Comparison A1, both classified as medium, consistent with the statistically significant differences identified by the Mann-Whitney U tests. In Comparison A2, all r_{rb} values remained below 0.27, confirming the distributional proximity between fixed-value mapping and the native metric scale across all PAQ attributes.

4.1.2. Comparisons B

Comparisons B1 to B4 assessed whether questionnaire configuration influenced ratings when both response formats were metric. Four bipolar adjective pairs (G4) were contrasted with eight single-attribute PAQs (G3) across four locations (P2–P5).

At the second location, P2 (Comparison B1, Table 5), Mann–Whitney U tests indicated statistically significant differences for *Eventful* ($U = 2127.50$, p-value$= 0.033$), *Pleasant* ($U = 1203.00$, p-value $= 0.004$), *Calm* ($U = 1074.50$, p-value < 0.001), *Uneventful* ($U = 1027.50$, p-value < 0.001), and *Monotonous* ($U = 555.00$, p-value < 0.001).

Specifically, the four bipolar adjective pairs configuration tended to yield higher ratings for *Pleasant*, as reflected in median values (66.00 vs. 50.00), *Calm* (56.00 vs. 42.00), *Uneventful* (40.00 vs. 24.00), and *Monotonous* (53.00 vs. 24.00), whereas *Eventful* tended toward higher values in the eight single-attribute PAQs (71.00 vs. 60.00). Differences in kurtosis also reflected these distributional contrasts, for example *Eventful* (1.00 in G3 vs. -0.90 in G4) and *Uneventful* (0.80 in G3 vs. -0.90 in G4).

At Comparison B2 (Table 6), Mann–Whitney U tests identified statistically significant configuration effects for *Uneventful* ($U = 1264.00$, p-value$=0.004$) and *Monotonous* ($U = 498.50$, p-value < 0.001). Consistent with these differences, the four bipolar adjective-pair format (G4) tended to produce higher ratings for both attributes (*Uneventful* median: 68.00 vs. 50.00; *Monotonous* median: 68.00 vs. 27.00), indicating stronger positioning along these perceptual dimensions under the bipolar configuration. Dispersion generally decreased in G4 for most PAQs, with only minor exceptions (e.g., *Chaotic*), suggesting a shift toward more consistent response patterns. Complementarily, distributional shape metrics indicated a greater concentration of high-end evaluations in G4, as reflected by more negative skewness in selected attributes (e.g., *Pleasant*: -1.40 vs. -0.70), reinforcing the tendency for responses to cluster toward higher ratings under the bipolar format at this location.

For Comparison B3 (Table 7), Mann–Whitney U tests revealed statistically significant differences for *Pleasant*

Table 3. Statistical data Comparison A1 — Group 1 transformed by Snell's method (G1_S): location P1, outdoor green area with vehicular traffic.

PAQ	U	p-value	r_{rb}	Mean (M) (G1_S|G2)	Median (Mdn) (G1_S|G2)	SD (G1_S|G2)	Skewness (g_1) (G1_S|G2)	Kurtosis (g_2) (G1_S|G2)
Eventful	1602.0	0.691	0.180	37.85|36.73	20.45|22.75	32.40|23.44	0.60|0.20	−1.08|−1.13
Vibrant	1146.0	0.003***	0.413	16.38|27.60	10.80|20.75	17.29|23.38	1.33|0.94	0.96|0.15
Pleasant	2379.5	< 0.001***	−0.219	87.36|74.17	91.38|76.50	12.89|18.73	−1.40|−0.73	1.14|0.12
Calm	1894.0	0.221	0.030	66.14|69.56	86.49|75.00	31.30|21.28	−0.45|−1.03	−1.34|0.75
Uneventful	1564.5	0.544	0.199	46.70|49.94	38.21|51.10	32.73|26.21	0.31|−0.26	−1.35|−0.97
Monotonous	1477.5	0.275	0.243	32.56|36.63	36.21|34.50	27.23|25.76	0.78|0.28	−0.37|−0.82
Chaotic	1127.0	0.002***	0.423	14.33|21.90	9.46|17.50	17.43|19.75	1.46|1.12	1.18|0.58
Annoying	1368.0	0.086*	0.299	16.54|22.65	6.96|17.00	20.93|21.87	1.09|0.69	−0.41|−0.59

*** p < 0.01; ** p < 0.05; * p < 0.10

Table 4. Statistical data Comparison A2 — Group 1 transformed by fixed-value mapping method (G1_i): location P1: outdoor green area with vehicular traffic.

PAQ	U	p-value	r_{rb}	Mean (M) (G1_i|G2)	Median (Mdn) (G1_i|G2)	SD (G1_i|G2)	Skewness (g_1) (G1_i|G2)	Kurtosis (g_2) (G1_i|G2)
Eventful	1771.0	0.591	0.093	40.30|36.70	25.00|33.80	29.80|23.40	0.30|0.20	−1.00|−1.10
Vibrant	1674.0	1.000	0.143	28.20|27.60	25.00|20.80	22.90|23.40	0.50|0.90	−0.50|0.20
Pleasant	1628.0	0.798	0.166	73.40|74.20	75.00|76.50	19.10|18.70	−0.60|−0.70	0.30|0.10
Calm	1434.5	0.182	0.265	64.10|69.60	75.00|75.00	27.40|21.30	−0.40|−1.00	−0.60|0.80
Uneventful	1589.0	0.638	0.186	47.60|40.90	50.00|51.50	29.20|26.20	0.10|−0.30	−1.00|−1.00
Monotonous	1727.5	0.767	0.115	38.30|36.60	50.00|34.50	26.30|25.80	0.10|0.30	−0.90|−0.80
Chaotic	1859.0	0.300	0.048	25.80|21.90	25.00|17.50	23.90|19.80	0.60|1.10	−0.60|0.60
Annoying	1823.5	0.402	0.066	29.40|22.60	25.00|17.00	28.10|21.90	0.50|0.70	−1.10|−0.60

*** p < 0.01; ** p < 0.05; * p < 0.10

Table 5. Statistical data Comparison B1: location P2, reverberant entrance hall with intense pedestrian circulation.

PAQ	U	p-value	r_{rb}	Mean (M) (G3|G4)	Median (Mdn) (G3|G4)	SD (G3|G4)	Skewness (g_1) (G3|G4)	Kurtosis (g_2) (G3|G4)
Eventful	2127.5	0.033**	−0.182	63.20|57.20	71.00|60.00	20.70|19.60	−1.30|−0.20	1.00|−0.90
Vibrant	1756.5	0.899	0.024	47.60|46.20	50.00|47.00	23.90|21.80	0.00|−0.10	−1.00|−0.50
Pleasant	1203.0	0.004***	0.331	51.50|63.50	50.00|66.00	22.70|21.40	−0.20|−0.40	−0.60|−0.00
Calm	1074.5	< 0.001***	0.403	43.10|59.30	42.00|56.00	23.80|21.60	0.20|0.00	−1.00|−0.70
Uneventful	1027.5	< 0.001***	0.429	29.00|42.80	24.00|40.00	19.70|19.60	0.90|0.20	0.80|−0.90
Monotonous	555.0	< 0.001***	0.692	25.60|53.80	24.00|53.00	18.30|21.80	0.70|0.10	0.10|−0.50
Chaotic	2006.5	0.104	−0.115	45.80|40.70	50.00|44.00	24.20|21.60	−0.30|0.00	−1.00|−0.70
Annoying	1904.0	0.356	−0.058	41.70|36.50	39.00|34.00	26.30|21.40	0.20|0.40	−0.90|0.00

*** p < 0.01; ** p < 0.05; * p < 0.10

($U = 1392.50$, p-value$=0.033$), *Uneventful* ($U = 1198.00$, p-value $= 0.002$), and *Monotonous* ($U = 680.00$, p-value < 0.001). Consistent with previous comparisons, median values were higher in the four bipolar adjective pairs (G4) for all three attributes: *Pleasant* (78.00 vs. 73.00), *Uneventful* (61.00 vs. 38.00), and *Monotonous* (65.00 vs. 33.00). Dispersion values were slightly lower in G4 for all PAQs, suggesting more stable response distributions under this configuration.

At the final location, P5 (Comparison B4, Table 8), Mann–Whitney U tests revealed statistically significant differences for *Calm* ($U = 1411.00$, p-value $= 0.022$), *Uneventful* ($U = 1188.00$, p-value < 0.001), *Monotonous* ($U = 813.50$, p-value < 0.001), and *Annoying* ($U = 1275.50$, p-value$=0.003$). The four bipolar adjective pairs configuration (G4) tended to produce higher ratings for *Calm* (85.00 vs. 80.00), *Uneventful* (84.00 vs. 73.00), *Monotonous* (85.00 vs. 59.00), and *Annoying* (28.00 vs. 18.00). Kurtosis values suggested more peaked distributions in G4 for some attributes, for example, *Calm* (2.90 vs. 0.40), *Uneventful* (1.20 vs. −0.40), and *Monotonous* (2.60 vs. −0.90).

Across all locations, *Uneventful* and *Monotonous* most consistently exhibited significant differences between questionnaire configurations, whereas *Vibrant* and *Chaotic* showed no significant variation under any location.

Effect sizes for Comparisons B reveal a consistent pattern across all four locations. For these comparisons, n_1 corresponds to G3 (eight single-attribute PAQs) and n_2 to G4 (four bipolar adjective pairs), such that positive r_{rb} values indicate that G4 tends to produce higher ratings for a given attribute, while negative values indicate the opposite. *Monotonous* exhibited the largest effect sizes across all locations (B1: $r_{rb} = 0.692$; B2: $r_{rb} = 0.723$; B3: $r_{rb} = 0.622$; B4: $r_{rb} = 0.548$), all classified as large, reflecting the strong and consistent tendency of the bipolar format to concentrate responses toward higher ratings along this dimension. *Uneventful* showed medium effects across all locations (B1: $r_{rb} = 0.429$; B2: $r_{rb} = 0.298$; B3:

Table 6. Statistical data Comparison B2: location P3, semi-open transitional area.

PAQ	**U**	p**-value**	r_{rb}	**Mean** (M) **(G3|G4)**	**Median (Mdn) (G3|G4)**	**SD (G3|G4)**	**Skewness** (g_1) **(G3|G4)**	**Kurtosis** (g_2) **(G3|G4)**
Eventful	1947.0	0.511	−0.082	39.50|36.30	33.50|32.00	24.90|23.50	0.50|0.70	−0.80|0.00
Vibrant	1957.5	0.476	−0.088	39.00|35.10	32.50|32.00	24.20|19.80	0.40|0.70	−0.80|0.80
Pleasant	1892.5	0.708	−0.052	78.40|77.40	81.00|81.00	17.10|17.20	−0.70|−1.40	−0.20|2.00
Calm	1838.0	0.927	−0.021	75.80|76.20	80.50|81.00	20.60|18.10	−1.00|−1.50	0.70|2.40
Uneventful	1264.0	0.004***	0.298	50.50|63.70	50.00|68.00	25.00|23.50	−0.20|−0.70	−1.20|0.00
Monotonous	498.5	< 0.001***	0.723	31.70|64.90	27.00|68.00	21.90|19.80	0.40|−0.70	−1.00|0.80
Chaotic	1574.0	0.201	0.125	20.10|23.80	19.50|19.00	17.60|18.10	0.70|1.50	−0.40|2.40
Annoying	1508.5	0.105	0.162	18.80|22.60	15.50|19.00	18.50|17.20	1.30|1.40	1.60|2.00

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$

Table 7. Statistical data Comparison B3: location P4, outdoor green area with distant traffic sounds.

PAQ	**U**	p**-value**	r_{rb}	**Mean** (M) **(G3|G4)**	**Median (Mdn) (G3|G4)**	**SD (G3|G4)**	**Skewness** (g_1) **(G3|G4)**	**Kurtosis** (g_2) **(G3|G4)**
Eventful	1975.5	0.357	−0.098	48.20|43.80	50.00|39.00	24.00|21.50	−0.10|0.20	−1.00|−0.90
Vibrant	1731.5	0.723	0.038	38.50|38.90	38.00|35.00	25.20|18.90	0.30|0.50	−0.90|−0.30
Pleasant	1392.5	0.033**	0.226	68.00|75.60	73.00|78.00	19.80|16.90	−0.40|−0.70	−0.50|0.10
Calm	1468.0	0.082*	0.184	63.20|70.60	67.00|75.00	24.90|19.90	−0.40|−0.90	−0.40|0.40
Uneventful	1198.0	0.002***	0.334	42.40|56.20	38.00|61.00	23.10|21.50	0.10|−0.20	−1.30|−0.90
Monotonous	680.0	< 0.001***	0.622	34.30|61.10	33.00|65.00	22.10|18.90	0.40|−0.50	−0.90|−0.30
Chaotic	1698.5	0.597	0.056	28.70|29.40	25.00|25.00	23.40|19.90	0.60|0.90	−0.70|0.40
Annoying	1789.0	0.958	0.006	24.90|24.40	22.00|22.00	19.00|16.90	0.70|0.70	−0.20|0.10

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$

Table 8. Statistical data Comparison B4: location P5, indoor classroom.

PAQ	**U**	p**-value**	r_{rb}	**Mean** (M) **(G3|G4)**	**Median (Mdn) (G3|G4)**	**SD (G3|G4)**	**Skewness** (g_1) **(G3|G4)**	**Kurtosis** (g_2) **(G3|G4)**
Eventful	2191.5	0.088*	−0.218	27.80|21.30	23.00|16.00	22.80|20.60	0.80|1.30	−0.10|1.20
Vibrant	1966.0	0.583	−0.093	22.60|19.50	17.00|15.00	21.00|18.20	1.10|1.60	0.70|2.60
Pleasant	2098.0	0.220	−0.166	73.40|68.40	79.00|72.00	19.60|22.40	−0.80|−0.80	0.20|0.10
Calm	1411.0	0.022**	0.216	74.30|82.90	80.00|85.00	21.20|15.60	−1.00|−1.50	0.40|2.90
Uneventful	1188.0	< 0.001***	0.340	64.60|78.70	73.00|84.00	25.60|20.60	−0.60|−1.30	−0.40|1.20
Monotonous	813.5	< 0.001***	0.548	55.50|80.50	59.00|85.00	27.40|18.20	−0.30|−1.60	−0.90|2.60
Chaotic	1918.0	0.762	−0.066	19.00|17.10	12.00|15.00	17.80|15.60	1.10|1.50	0.80|2.90
Annoying	1275.5	0.003***	0.291	20.10|31.60	18.00|28.00	17.90|22.40	0.90|0.80	0.00|0.10

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$

$r_{rb} = 0.334$; B4: $r_{rb} = 0.340$), while the remaining attributes showed small to negligible effects, with localized exceptions for *Calm*, *Pleasant*, and *Annoying* at specific locations. In particular, *Eventful* presented negative r_{rb} values across all four locations (B1: $r_{rb} = -0.182$; B2: $r_{rb} = -0.082$; B3: $r_{rb} = -0.098$; B4: $r_{rb} = -0.218$), indicating that the eight single-attribute configuration consistently captured higher eventfulness ratings regardless of location. This ordering confirms that the bipolar four-adjective pair configuration most strongly differentiates from the eight single-attribute format along the Eventfulness–Uneventfulness perceptual axis.

4.2. Two-dimensional perceptual soundscape representations

Two-dimensional perceptual soundscape graphs were constructed in accordance with ISO/TS 12913-3, plotting participant responses along the *Pleasantness* (x-axis) and *Eventfulness* (y-axis) dimensions. Kernel Density Estimation (KDE) contours represent areas of higher response concentration, while the four quadrants correspond to the perceptual categories defined by the standard: *Vibrant* (upper-right), *Calm* (lower-right), *Monotonous* (lower-left), and *Chaotic* (upper-left).

4.2.1. Comparisons A

Figure 3 shows the graphs for comparisons A. For Group 1, two conversion methods were applied to the original ordinal PAQ scale: Snell conversion (G1_S) and fixed-value mapping (G1_i). The KDE distribution for G1_S shows response density concentrated within the *Calm* quadrant, with a moderate spread towards *Monotonous* and *Vibrant*, and minimal density in *Chaotic*. The centroid position lies in the positive *Pleasantness* region and negative *Eventfulness* region, indicating that judgments remained predominantly *Calm*.

Under the fixed-value mapping (Figure 3, G1_i), responses display greater dispersion along the *Eventfulness* axis, extending into the *Vibrant* quadrant while maintaining maximum density in *Calm*.

In Group 2, which uses the native metric scale, KDE contours are more evenly distributed between *Calm* and

Vibrant, with the peak density shifted upward on the *Eventfulness* axis compared with both G1_S and G1_i. Figure 3 (G2) shows that the *Monotonous* and *Chaotic* quadrants remain sparsely populated, indicating low frequency of negative *Pleasantness* ratings across all formats.

4.2.2. Comparisons B

When comparing Groups 3 and 4, distinct distribution patterns were observed across the four locations.

Figure 4, corresponding to Location 2, shows that both configurations concentrated responses in the *Vibrant* quadrant. G3_1 exhibited a broader spread extending into the *Calm* and *Chaotic* regions, whereas G4_1 formed a more compact cluster with its centroid positioned higher on the *Eventfulness* axis. Figure 5, corresponding to Location 3, show responses shifted predominantly toward the *Calm* quadrant. G3_2 displayed a dense, centrally located cluster between *Pleasant* and *Calm*, while G4_2 displayed a cluster between *Calm* and *Uneventful*.

Locations 4 and 5 (Figures 6 and 7) also exhibited predominantly *calm* perceptual profiles.

In Location 4, G3_3 spanned the *Calm* and *Vibrant* quadrant, presenting an elongated contour along the *Pleasantness* axis while G4_3 maintained a compact, high-density cluster centered in *Calm*, positioned between *Pleasant* and *Uneventful*.

At Location 5, both groups concentrated strongly in the *Calm* quadrant, near the transition towards *Uneventful*. G3_4 displayed a secondary density peak extending toward *Monotonous*, whereas G4_4 retained a single, well-defined contour. In both cases, the *Chaotic* quadrant was nearly absent, indicating a consistent perception of positively valanced and low-*eventfulness* soundscapes.

5. Discussion

This study employed a comparative design to evaluate how different conversion methods (Snell vs. fixed-value mapping) and distinct PAQ questionnaire configurations (four bipolar adjective pairs vs. eight single attributes) affect data interpretation. By integrating inferential statistics, descriptive parameters, and perceptual representations in accordance with ISO/TS 12913-3 [11], the approach provides empirical evidence of the methodological consequences associated with these analytical choices.

5.1. Effects of the response scale transformation – Comparison A

The conversion of ordinal responses into a metric format produced distinct outcomes depending on the method applied. Using Snell's conversion (G1_S), Mann–Whitney U tests indicated statistically significant differences for the attributes *Vibrant*, *Pleasant*, and *Chaotic*. Descriptive statistics revealed systematic redistributions in median, skewness, and kurtosis, indicating changes in central tendency, distributional shape, and concentration. For instance, *Pleasant* presented a higher median, reduced variability, and increased kurtosis under Snell's conversion, resulting in more compact KDE contours in the *Calm* quadrant compared with the metric scale. Conversely, *Vibrant* and *Chaotic* showed lower central tendency values, consistent with reduced density in high-*Eventfulness* areas of the ISO/TS 12913-3 [11] graphical representation.

The mechanism underlying these variations derives from Snell's empirical cumulative probability-based adjustment, which repositions category thresholds according to the observed distribution, approximating them on a continuous latent scale, thereby mitigating distortions introduced by arbitrary fixed-interval mappings [35, 40–42]. However, because Snell's transformation estimates item-specific thresholds from the observed response distribution, it can affect some PAQs more than others. This effect tends to be stronger when responses occupy a restricted portion of the circumplex (e.g., concentrated in *Calm* with sparse ratings in *Chaotic*), where asymmetric or truncated distributions amplify item-dependent rescaling. As a result, downstream comparisons and models that are sensitive to distributional shape or relative positioning may be influenced by these attribute-specific redistributions.

In contrast, the fixed-interval mapping (0, 25, 50, 75, 100) applied in Comparison A2 did not produce statistically significant differences compared to the metric scale across any PAQ. Descriptive values were closely aligned between G1_i and G2, and their KDE contours exhibited extensive overlap, with only minor variations along the *Eventfulness* axis. This pattern indicates that fixed-value mapping effectively preserves rank order and imposes an interval-like structure without substantially altering distributional properties, a point also noted by Menold [34] in relation to scale polarity and granularity. Nevertheless, as no perceptual recalibration occurs, subtle compressions or expansions may remain undetected.

In summary, Comparisons A show that Snell's conversion modified the response distribution in an attribute-dependent manner, whereas fixed-value mapping produced outcomes more closely aligned with the native metric scale (0-100) (Figure 8). For statistical procedures sensitive to distributional shape, such as parametric regressions under their required assumptions or percentile-based analyses, these discrepancies can be consequential and should be documented.

5.2. Effects of questionnaire structure – Comparison B

When both groups used exclusively numeric responses (G3 and G4), observed differences stemmed mainly from the PAQ questionnaire configuration. Across Comparisons B1 to B4 and all four locations (P2 to P5), *Monotonous* and *Uneventful* most frequently showed statistically significant differences, whereas no statistically significant differences were detected for *Vibrant* and *Chaotic*. In the circumplex graphs, the configuration with four bipolar adjective pairs (G4) consistently produced

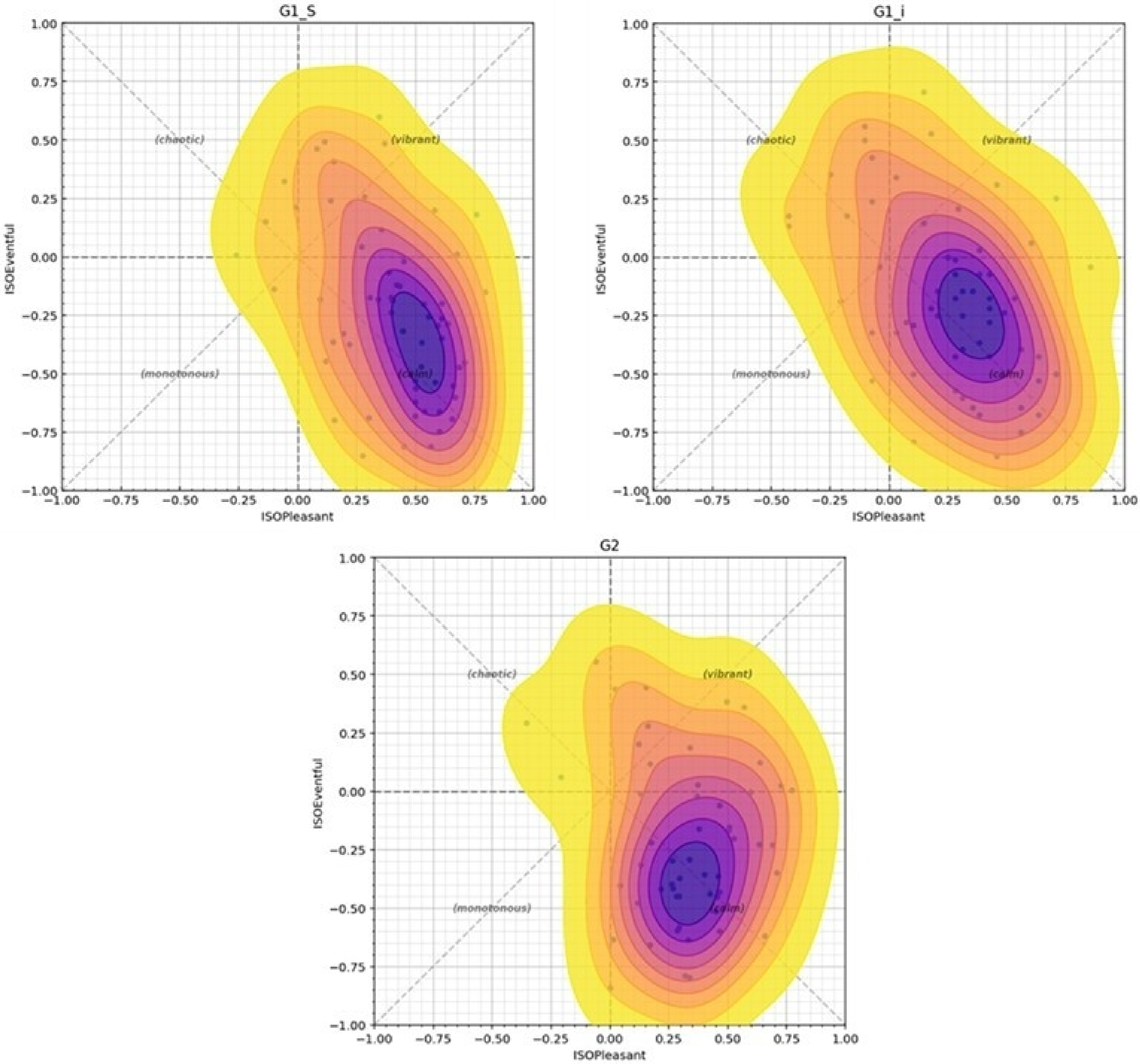

Figure 3. Soundscape circumplex graphs: Snell conversion (G1_S), fixed-value mapping (G1_i), and data collected in metric scale (G2) for Point P1 (open green area).

more compact density clusters in the *Calm* quadrant, whereas the eight single-attribute PAQs (G3) exhibited KDE contours extending toward *Vibrant* and *Chaotic* (Figure 9).

At Location 2, both groups occupied the positive *Pleasantness* axis, but G3_1 extended further into *Vibrant* and *Chaotic*, while G4_1 remained concentrated near *Calm*, showing lower *Eventfulness*. Median values followed this pattern, with higher *Eventful* scores in G3 and higher *Calm* and *Pleasant* scores in G4.

At Location 3, distributions shifted toward *Calm*: G3_2 formed a density centroid, whereas G4_2 extended slightly toward *Monotonous*. Statistically significant differences occurred for *Uneventful* and *Monotonous*, consistent with the reduced spread observed in G4.

At Location 4, both covered *Calm* and part of *Vibrant*, although G3_3 presented an elongated contour along the Pleasantness axis, while G4_3 remained compact near the *Calm–Vibrant* interface. Statistically significant differences were found for *Pleasant*, *Uneventful*, and *Monotonous*, aligning with the stronger concentration observed in G4.

For Location 5, both groups clustered in the *Calm* quadrant. G3_4 exhibited a secondary density peak toward *Monotonous*, whereas G4_4 maintained a single, sharply defined centroid. Descriptive kurtosis was higher

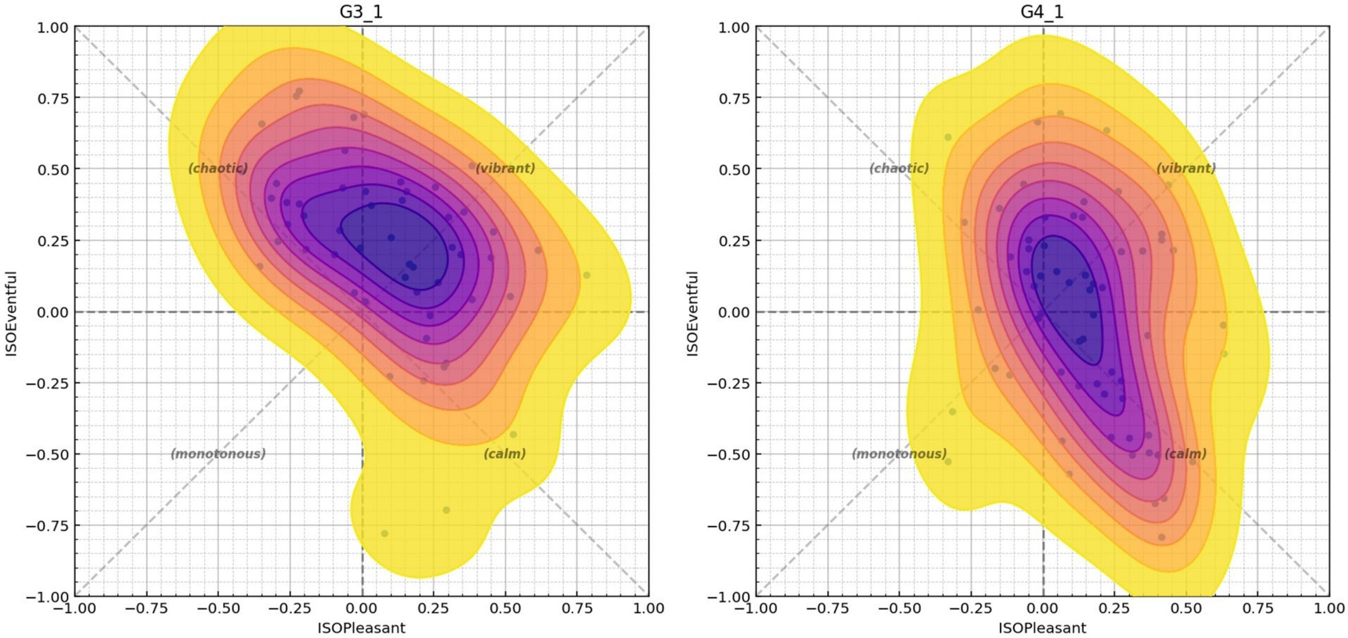

Figure 4. Soundscape circumplex graphs for Point 2 (P2): G3_1 (eight single-attribute PAQs) and G4_1 (four bipolar adjective pairs) at point P2 (entrance hall).

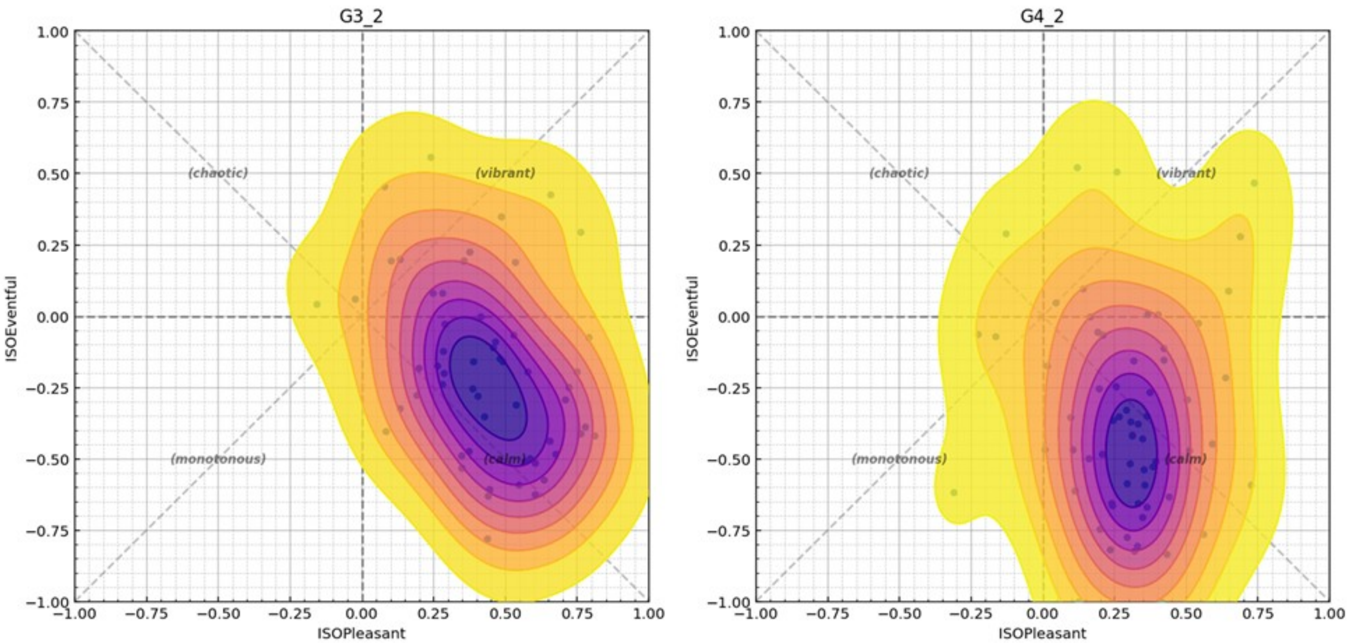

Figure 5. Soundscape circumplex graphs for Point 3 (P3): G3_2 (eight single-attribute PAQs) and G4_2 (four bipolar adjective pairs) at point P3 (transitional area).

in G4 for *Calm* and *Monotonous*, reflecting more peaked response distributions.

Across all locations, from the highly *Eventful* P2 to the *Uneventful* P5, these outcomes show that the reduced bipolar configuration concentrates responses and limits semantic overlap. This tendency is consistent with findings from standardized environmental noise questionnaires such as those developed by ICBEN [30]. In contrast, the expanded eight-attribute configuration captures a broader perceptual range, enhancing spatial granularity but introducing greater response variability.

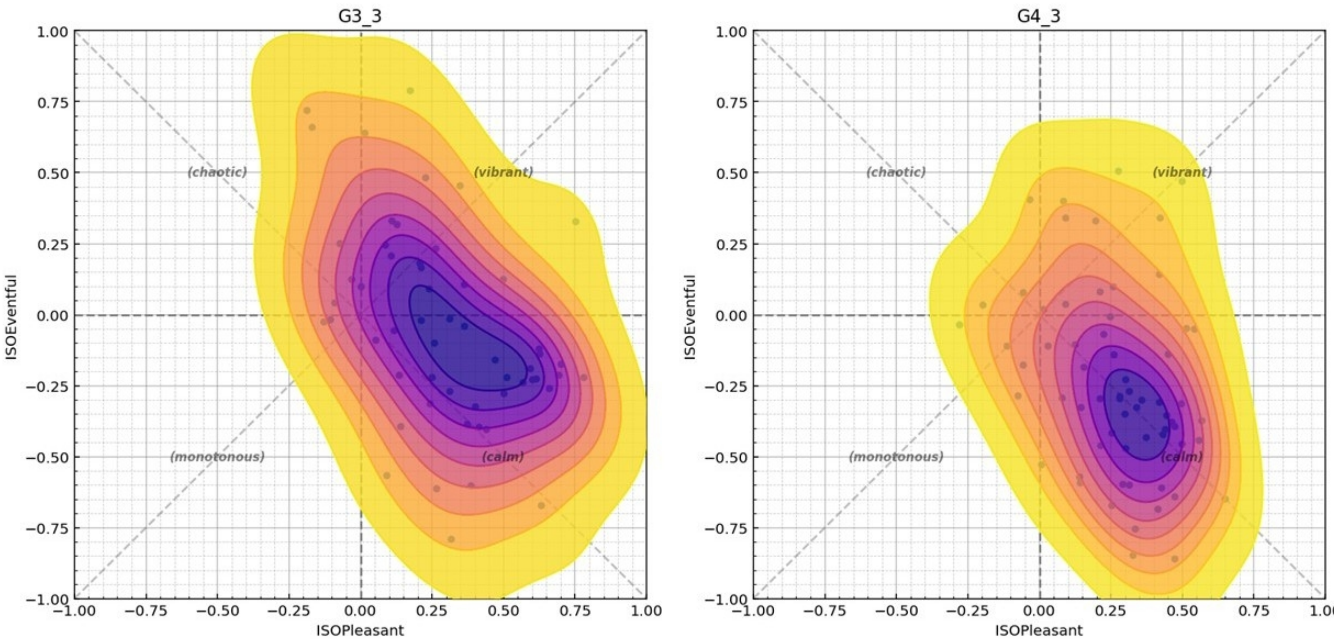

Figure 6. Soundscape circumplex graphs for Point 4 (P4): G3_3 (eight single-attribute PAQs) and G4_3 (four bipolar adjective pairs) at point P (open green area).

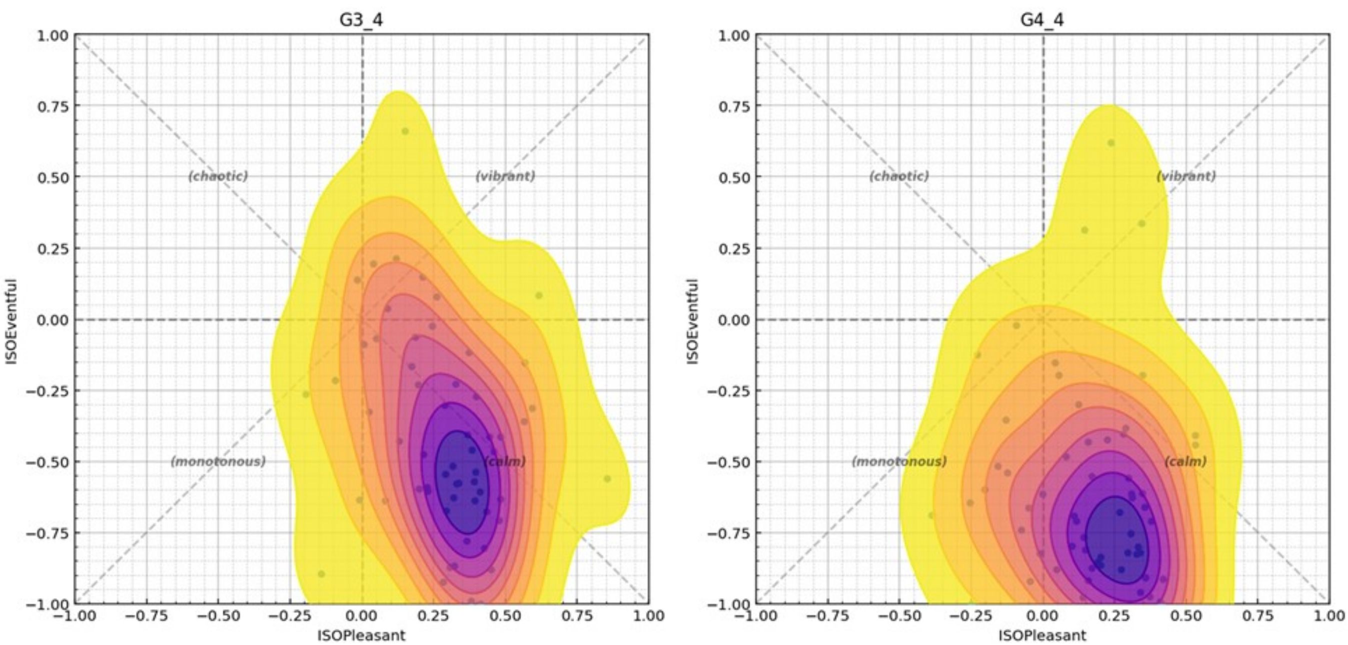

Figure 7. Soundscape circumplex graphs for Point 5 (P5): G3_4 (eight single-attribute PAQs) and G4_4 (four bipolar adjective pairs) at point P5 (classroom).

5.3. Convergence between statistical and perceptual analyses

A key strength of this dataset lies in the convergence between inferential outcomes and perceptual representations. Attributes showing significant Mann–Whitney U tests differences were consistently associated with perceptual shifts in KDE graphs, including centroid relocation, contour deformation, and density redistribution across quadrants.

In Comparison A1, attribute-specific redistributions under Snell's transformation were clearly visible through contour compression or expansion, whereas in Comparison A2, statistical similarity was reflected by nearly overlapping KDE profiles.

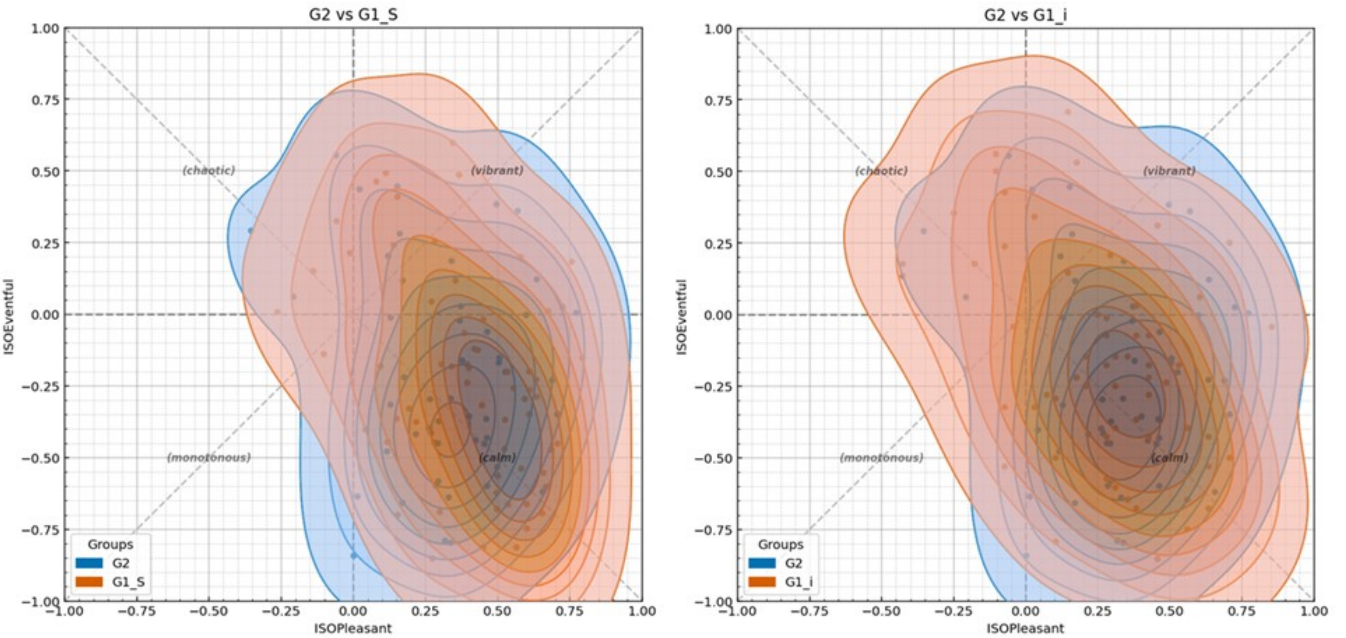

Figure 8. Comparative soundscape circumplex graphs between Group 2 (metric scale), in blue, and Group 1 conversed using Snell's method (left) or fixed-value (right), both in orange, for Point P1 (open green area).

Across the comparisons B1-B4, attributes with high statistical sensitivity (*Monotonous*, *Uneventful*) also showed broader tails in the eight single-attribute configuration (G3) and higher kurtosis with tighter clustering in the configuration with four bipolar adjective pairs (G4). This alignment between numerical and perceptual evidence reinforces the reliability of the findings, as the observed patterns emerge consistently across independent analytical dimensions. The inclusion of skewness and kurtosis metrics further clarified when perceptual evaluations were directionally biased yet distributionally homogeneous.

5.4. Methodological and practical implications

Three main methodological implications arise:

1. Conversion of ordinal responses: To harmonize ordinal and numeric formats, the fixed-interval mapping approach proved to be more stable in preserving observed distributional properties. Snell's conversion, based on latent threshold modeling, offers perceptual calibration by adapting to empirical distributions, but it can also introduce attribute-specific redistributions. When applied, sensitivity analyses or transparent supplemental reporting are recommended to ensure interpretability.
2. Configuration of the PAQ questionnaire: The four bipolar adjective-pair format yielded consistently tighter circumplex clusters, whereas the eight single-attribute version reflected a wider spread of perceptual responses. Accordingly, configuration choice should follow the intended use: bipolar pairs support focused, quadrant-based profiling, while the expanded attribute set is more appropriate for exploratory mapping of perceptual heterogeneity.
3. Integration of statistical and perceptual tools: Soundscape circumplex graphs enhanced with KDE are not merely illustrative; they make distributional structure visible (e.g., multimodality, directional elongation, and clustering) in ways that summary statistics alone cannot capture. In this study, these distributional signatures were consistent with the interaction between questionnaire configuration and location-specific context: acoustically heterogeneous settings with multiple concurrent sources (e.g., P1: low traffic, pedestrians, and biophonic sounds; P2: a reverberant entrance hall with intense circulation and partial exposure to external noise; P3: a semi-open transitional space near circulation/parking; P4: human activities with distant traffic and natural sounds) tended to produce broader and more anisotropic KDE contours, whereas more source-limited indoor conditions (P5: indoor space dominated by human activity and room acoustics/reverberation) tended to concentrate responses. Taken together, combining non-parametric inference with KDE-based circumplex mapping and a minimal description of site characteristics establishes a more rigorous and context-sensitive framework for interpreting PAQ-based soundscape outcomes.
4. Socio-cultural and cross-location generalization: The effects observed here were obtained with a homogeneous sample (architecture students, single campus, Brazilian Portuguese speakers), which represents a specific socio-cultural profile. Evidence from cross-cultural soundscape research [18, 19] suggests that attributes along the *Eventful–Uneventful* axis carry par-

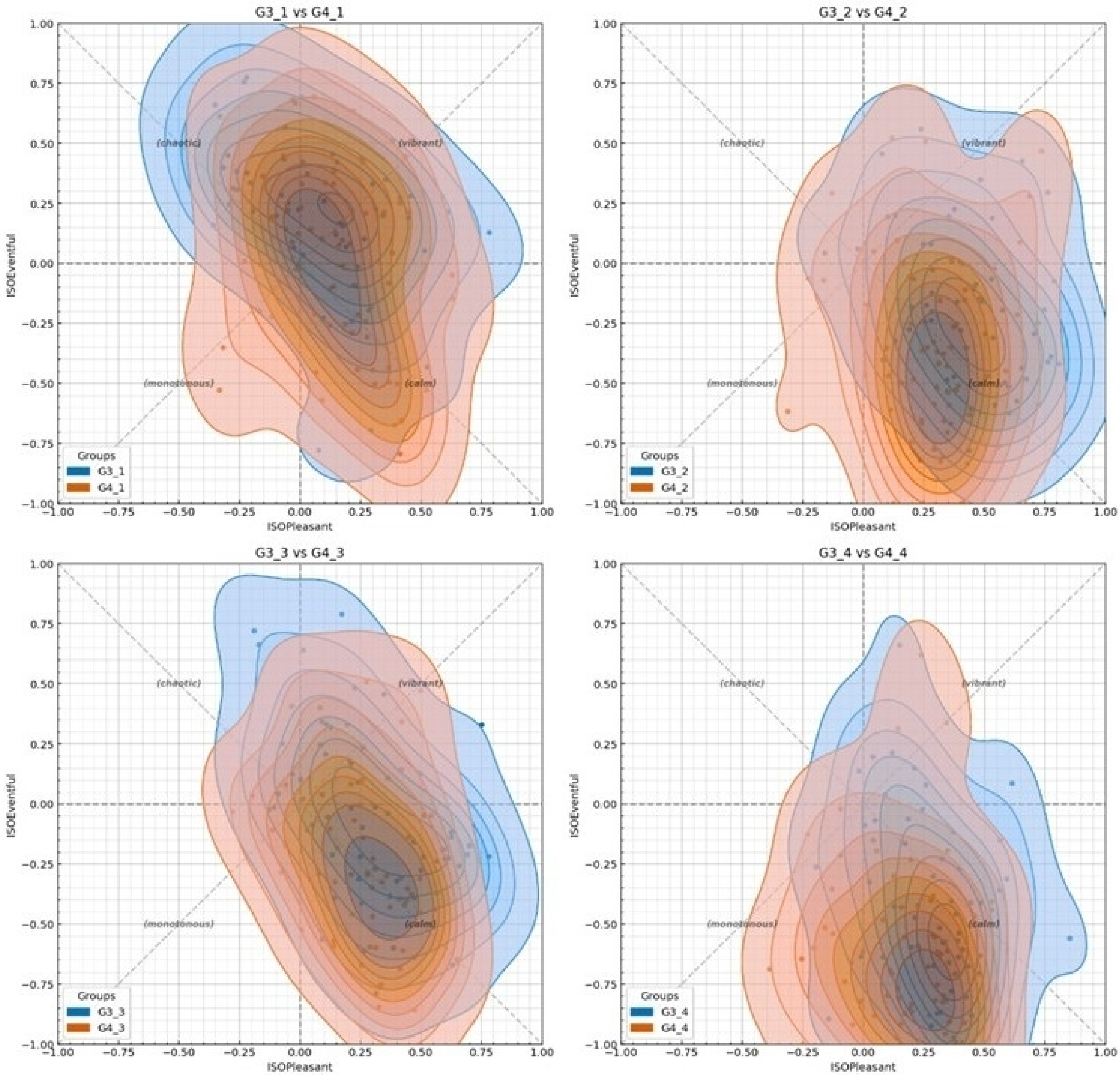

Figure 9. Comparative soundscape circumplex graphs for Group 3 (eight single-attributes), in blue, and Group 4 (four bipolar adjective pairs), in orange, across the four locations (P2 to P5).

ticular semantic fragility in Portuguese, and broader psychometric literature indicates that bipolar scale formats and response tendencies vary systematically across cultural groups [51–53]. Future applications should therefore test whether the configuration effects reported here are stable across populations with different linguistic backgrounds, acoustic familiarity, or urban experience. Incorporating measurement invariance testing across groups would be a concrete methodological improvement.

6. Conclusion

This study demonstrated that methodological decisions concerning both the conversion of ordinal responses and the configuration of PAQ questionnaires significantly influence the statistical and perceptual representation of soundscape evaluations. In this context, effectiveness is understood as the extent to which a given response format supports the intended analytical objective while maintaining sensitivity, distributional stability, and interpretative clarity. By integrating inferential statistics, descriptive parameters, and two-dimensional soundscape circumplex graphs, the analyses provided a multidimensional assessment that not only quantifies differences but also reveals their perceptual manifestations.

In the comparison between ordinal-to-metric conversions, Snell's method produced attribute-specific redistributions, altering medians, kurtosis, and perceptual density contours, whereas fixed-value mapping remained closely aligned with the native metric scale (0–100). These findings indicate that conversion methods are not interchangeable. Snell's method can enhance sensitivity to certain perceptual nuances but may also reshape response distributions, potentially affecting subsequent analyses. Fixed-value mapping, on the other hand, offers greater stability when the analytical objective is harmonization across formats.

Regarding questionnaire configuration, the bipolar four-adjective pair configuration tended to yield more compact density clusters, particularly at locations with lower acoustic complexity and reduced source heterogeneity (P3, P4, and P5), whereas the eight single-attribute configuration better captured perceptual variability in acoustically heterogeneous environments with multiple concurrent sources (P2). A systematic analysis of sound-source dominance as an independent variable, however, would require a dedicated experimental design and is suggested as a direction for future research. A salient finding concerns the attributes *Uneventful* and *Monotonous*, which showed the greatest distributional instability under methodological conditions. This variability cannot be explained solely by scale conversion. Previous studies [18–20] reported semantic ambiguity and low spontaneous usage of these attributes, indicating conceptual fragility of the *Eventful–Uneventful* axis. Recognizing this limitation is essential for adapting ISO 12913 PAQs to Brazilian Portuguese and improving the cross-cultural robustness of soundscape tools.

A key strength of this work lies in the convergence between statistics and perceptual soundscape circumplex graphs, reinforcing the validity of the observed effects. Attributes that differed significantly in non-parametric tests also showed corresponding shifts in centroid positions, contour shapes, and quadrant distribution in the circumplex graphs. The combined use of descriptive statistics (including kurtosis and skewness) and KDE-based visualizations allowed a richer interpretation of distributional characteristics beyond central tendency alone.

By reporting these results, the study advances methodological transparency, explicitly demonstrating how variations in response scale format and PAQ questionnaire configuration influence both statistical outcomes and circumplex-based perceptual representations in soundscape analysis. On this basis, it provides empirically grounded recommendations for selecting conversion methods and questionnaire designs according to analytical objectives, whether prioritizing perceptual sensitivity, distributional stability, or cross-study comparability.

Ultimately, these findings contribute to the refinement of ISO/TS 12913-3 [11] applications, offering a replicable analytical framework that integrates psychometric rigor with perceptual mapping. Such an approach enhances the interpretative power of soundscape analyses and may contribute to more informed, evidence-based decisions in urban acoustic planning and environmental quality assessment.

By demonstrating how specific questionnaire configurations and scale conversions affect both statistical outcomes and perceptual circumplex graphs, the study provides a transferable framework for optimizing public surveys, refining noise mapping practices, and integrating perceptual indicators into environmental quality standards. This alignment between methodological precision and practical applicability points to the broader potential of PAQ-based instruments to contribute to soundscape assessment in urban planning contexts, as discussed by Michalski *et al.* [20] in the Brazilian standardization framework. Since the observed trends were obtained from a sample within a specific sociocultural and location context, the generalizability of these results to other contexts is necessarily limited. The effects associated with questionnaire configuration and the distributional stability of fixed-value mapping relative to Snell's conversion were consistent across all comparison sets, supporting the internal validity of the results. Replication studies employing diverse population profiles, urban typologies, and linguistic backgrounds are nonetheless needed to determine the extent to which the effects are stable across broader contexts. Incorporating formal measurement invariance testing between questionnaire formats would represent a concrete and recommended methodological advance in this direction.

Finally, the study provides consistent evidence within the scope investigated, offering a solid basis for comparing response formats and questionnaire structures. Nevertheless, some limitations should be acknowledged when interpreting the reach of these results. The sample consisted of university students, and data collection was conducted within the urban context of a university campus; therefore, extrapolation to more diverse sociodemographic profiles and to other urban typologies should be undertaken with due caution. In addition, formal measurement invariance was not examined at this stage. Future research can strengthen and extend these conclusions by replicating the protocol across different population groups and urban contexts, explicitly testing invariance between formats and questionnaire structures, and integrating these findings with objective acoustic indicators, thereby advancing more comprehensive and multimodal approaches to soundscape assessment.

Funding

This study was partially supported by grants number #2021/14221-5 and #2021/02915-2 from the São Paulo Research Foundation (FAPESP), and by grants number 16159/2021-2 and 308206/2025-8, and by project number 406603/2023-5 from the National Council for Scientific and Technological Development (Conselho Nacional de Desenvolvimento Científico e Tecnológico, CNPq – Brazil).

Conflicts of interest

The authors declare they have nothing to disclose.

Data availability statement

The datasets generated and analyzed during the current study are available from the corresponding authors upon reasonable request.

Author contribution statement

Lucas Rafael Ferreira: conceptualization, data curation, formal analysis, investigation, methodology, validation, visualization, writing – original draft, writing – review & editing. **Leonardo Marques Monteiro**: project administration, supervision, writing – review & editing. **Nara Gabriela de Mesquita Peixoto**: conceptualization, data curation, formal analysis, investigation, methodology, validation, visualization, writing – original draft, writing – review & editing. **Ranny Loureiro Xavier Nascimento Michalski**: conceptualization, data curation, formal analysis, investigation, methodology, validation, visualization, writing – original draft, writing – review & editing..

Ethical approval

This study was approved by the Research Ethics Committee (CEP) of the University of São Paulo (CAEE 80357224.1.0000.5390 and 63954322.4.0000.5390).

References

1. Panagopoulos T, González Duque JA, Bostenaru Dan M (2016) Urban planning with respect to environmental quality and human well-being, Environ Pollut, 208, 137–144. https://doi.org/10.1016/j.envpol.2015.07.038
2. van Kamp I, Leidelmeijer K, Marsman G, de Hollander A (2003) Urban environmental quality and human well-being, Landsc Urban Plan, 65, 5–18. https://doi.org/10.1016/S0169-2046(02)00232-3
3. Schuck M, Haselhoff T, Lawrence BT, Moebus S, Fiebig A (2025) Advancing the characterization of urban acoustic environments through multimetric analysis, Acta Acustica, 9, 66. https://doi.org/10.1051/aacus/2025051
4. Brooks BM, Schulte-Fortkamp B, Woolworth D s. (2024) Soundscape in application, J Acoust Soc Am, 156, A93–A93. https://doi.org/10.1121/10.0035218
5. Davies WJ, Adams MD, Bruce NS, Cain R, Carlyle A, Cusack P, et al. (2013) Perception of soundscapes: An interdisciplinary approach, Appl Acoust, 74, 224–231. https://doi.org/10.1016/j.apacoust.2012.05.010
6. Schulte-Fortkamp B, Brooks BM (2024) Applying soundscape: A matter of participation, J Acoust Soc Am, 155, A113–A113. https://doi.org/10.1121/10.0026993
7. Durbridge S, Murphy DT (2023) Assessment of soundscapes using self-report and physiological measures, Acta Acustica, 7, 6. https://doi.org/10.1051/aacus/2022059
8. Ma KW, Mak CM, Wong HM (2021) Effects of environmental sound quality on soundscape preference in a public urban space, Appl Acoust, 171, 107570. https://doi.org/10.1016/j.apacoust.2020.107570
9. International Organization for Standardization (2014) ISO 12913-1 Acoustics — Soundscape — Part 1: Definition and conceptual framework.
10. International Organization for Standardization (2018) ISO/TS 12913-2 Acoustics — Soundscape — Part 2: Data collection and reporting requirements.
11. International Organization for Standardization (2025) ISO/TS 12913-3 Acoustics — Soundscape — Part 3: Data analysis.
12. Holmes A, Neal M, Zahorik P (2023) Loudness in rooms and the sound strength parameter: Linking perception and physical acoustics, J Acoust Soc Am, 153, A156–A156. https://doi.org/10.1121/10.0018487
13. Chan T-C, Wu B-S, Lee Y-T, Lee P-H, Deng R-H (2023) Exploring relationships among soundscape perception, spatiotemporal sound characteristics, and personal traits through social media, Noise Mapp, 10. https://doi.org/10.1515/noise-2022-0174
14. Wen X, Meng Q, Yin Y, Yang D, Li M, Kang J (2025) Exploring the boundaries of indoor combined thermal-acoustic environmental effects on comfort perceptions, Build Environ, 271, 112580. https://doi.org/10.1016/j.buildenv.2025.112580
15. Torresin S, Albatici R, Aletta F, Babich F, Oberman T, Siboni S, et al. (2020) Indoor soundscape assessment: A principal components model of acoustic perception in residential buildings, Build Environ, 182, 107152. https://doi.org/10.1016/j.buildenv.2020.107152
16. Aletta F, Kang J, Axelsson Ö (2016) Soundscape descriptors and a conceptual framework for developing predictive soundscape models, Landsc Urban Plan, 149, 65–74. https://doi.org/10.1016/j.landurbplan.2016.02.001
17. Nagahata K, Saito M, Funayama T, Ishii T (2023) Can the perceived affective quality scales evaluate the restorativeness of soundscapes? INTER-NOISE NOISE-CON Congr Conf Proc, 268, 3659–3666. https://doi.org/10.3397/IN_2023_0520
18. Monteiro Antunes S, Nascimento Michalski RLX, Luiza de Ulhôa Carvalho M, Alves S, Cláudio Ribeiro L (2023) A European and Brazilian cross-national investigation into the Portuguese translation of soundscape perceptual attributes within the SATP project, Appl Acoust, 211, 109472. https://doi.org/10.1016/j.apacoust.2023.109472
19. de Mesquita Peixoto NG, Rafael Ferreira L, Afonso RC, Comissoli VG, Marques Monteiro L (2024) A methodological approach for soundscape attributes translation using Twitter data: A case study in the Portuguese language, Appl Acoust, 218, 109880. https://doi.org/10.1016/j.apacoust.2024.109880
20. Michalski, RLXN, Melo, VSGd, Engel, MS (2026) Soundscape Standardization and Sustainability: An Evaluation of ISO 12913 in the Brazilian Context, Sustainability, 18(6), 2752. https://doi.org/10.3390/su18062752
21. Simms LJ, Zelazny K, Williams TF, Bernstein L (2019) Does the number of response options matter? Psychometric perspectives using personality questionnaire data, Psychol Assess, 31, 557–566. https://doi.org/10.1037/pas0000648
22. DeCastellarnau A (2018) A classification of response scale characteristics that affect data quality: a literature review, Qual Quant, 52, 1523–1559. https://doi.org/10.

1007/s11135-017-0533-4
23. Dawes J (2008) Do Data Characteristics Change According to the Number of Scale Points Used? An Experiment Using 5-Point, 7-Point and 10-Point Scales, Int J Mark Res, 50, 61–104. https://doi.org/10.1177/147078530805000106
24. Cabooter E, Weijters B, Geuens M, Vermeir I (2016) Scale format effects on response option interpretation and use, J Bus Res, 69, 2574–2584. https://doi.org/10.1016/j.jbusres.2015.10.138
25. Höhne JK, Lenzner T (2018) New insights on the cognitive processing of agree/disagree and item-specific questions. J Surv Stat Methodol 6(3):401–417. https://doi.org/10.1093/jssam/smx028
26. Tarlao C, Steele D, Guastavino C (2022) Assessing the ecological validity of soundscape reproduction in different laboratory settings. PLoS ONE 17(6):e0270401. https://doi.org/10.1371/journal.pone.0270401
27. Rachman Z, Aletta F, Kang J (2024) Exploring soundscape assessment methods in office environments: A systematic review. Buildings 14(11):3408. https://doi.org/10.3390/buildings14113408
28. Schreckenberg D, Belke C, Spilski J (2018) The Development of a Multiple-Item Annoyance Scale (MIAS) for Transportation Noise Annoyance, Int J Environ Res Public Health, 15, 971. https://doi.org/10.3390/ijerph15050971
29. International Organization for Standardization (2021) ISO/TS 15666 Acoustics — Assessment of noise annoyance by means of social and socio-acoustic surveys.
30. Brink M, Schreckenberg D, Vienneau D, Cajochen C, Wunderli J-M, Probst-Hensch N, et al. (2016) Effects of Scale, Question Location, Order of Response Alternatives, and Season on Self-Reported Noise Annoyance Using ICBEN Scales: A Field Experiment, Int J Environ Res Public Health, 13, 1163. https://doi.org/10.3390/ijerph13111163
31. Zhang J, Chen K, Li H, Chen X, Dong N (2023) The effects of rating scales and individual characteristics on perceived annoyance in laboratory listening tests, Appl Acoust, 202, 109137. https://doi.org/10.1016/j.apacoust.2022.109137
32. Arvidsson R (2019) On the use of ordinal scoring scales in social life cycle assessment, Int J Life Cycle Assess, 24, 604–606. https://doi.org/10.1007/s11367-018-1557-2
33. King LA, King DW, Klockars AJ (1983) Dichotomous and Multipoint Scales Using Bipolar Adjectives, Appl Psychol Meas, 7, 173–180. https://doi.org/10.1177/014662168300700205
34. Menold N (2023) Verbalization of Rating Scales Taking Account of Their Polarity, Field Methods, 35, 378–391. https://doi.org/10.1177/1525822X231151314
35. Snell EJ (1964) A Scaling Procedure for Ordered Categorical Data, Biometrics, 20, 592. https://doi.org/10.2307/2528498
36. Mitchell A, Aletta F, Kang J (2022) How to analyse and represent quantitative soundscape data, JASA Express Lett, 2. https://doi.org/10.1121/10.0009794
37. Harvard Humanitarian Initiative (n.d.) KoboToolbox platform.https://hhi.harvard.edu/kobotoolbox
38. R Core Team (2024) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.https://www.R-project.org/
39. JASP Team (2024) JASP.https://jasp-stats.org/previous-versions/
40. Maruf M (2023) Alternative approach to analysing data obtained with likert scale, Route Educ Soc Sci J, 10, 96–102. https://doi.org/10.17121/ressjournal.3439
41. Chakrabartty SN (2024) Port Sustainability Index: Methodological Issues, J Energy Environ Sci, 2, 1–10. https://doi.org/10.23880/jeesc-16000107
42. Prudjinski M, Hadar I, Luria G (2024) Exploring the Role of Team Security Climate in the Implementation of Security by Design: A Case Study in the Defense Sector, IEEE Trans Softw Eng, 50, 1065–1079. https://doi.org/10.1109/TSE.2024.3374114
43. Lindborg PM (2016) A taxonomy of sound sources in restaurants, Appl Acoust, 110, 297–310. https://doi.org/10.1016/j.apacoust.2016.03.032
44. Peixoto NGM, Rafael Ferreira L, Klein ME, Michalski RLXN, Monteiro LM (2023) Influence of soundscape on quality of work from home during the second phase of the pandemic in Brazil, Noise Mapp, 10. https://doi.org/10.1515/noise-2022-0175
45. Mann HB, Whitney DR (1947) On a Test of Whether One of Two Random Variables is Stochastically Larger than the Other, Ann Math Stat, 18(1), 50–60. https://doi.org/10.1214/aoms/1177730491
46. Nachar N (2008) The Mann–Whitney U: A Test for Assessing Whether Two Independent Samples Come from the Same Distribution, Tutorials Quant Methods Psychol, 4(1), 13–20. https://doi.org/10.20982/tqmp.04.1.p013
47. Divine GW, Norton HJ, Barón AE, Juarez-Colunga E (2018) The Wilcoxon–Mann–Whitney Procedure Fails as a Test of Medians, Am Stat, 72(3), 278–286. https://doi.org/10.1080/00031305.2017.1305291
48. Hart A (2001) Mann–Whitney test is not just a test of medians: differences in spread can be important, BMJ, 323(7309), 391–393. https://doi.org/10.1136/bmj.323.7309.391
49. Fay MP, Proschan MA (2010) Wilcoxon–Mann–Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules, Stat Surv, 4, 1–39. https://doi.org/10.1214/09-SS051
50. McKnight P, Najab J (2010) Mann–Whitney U Test, in: Salkind NJ (ed) The SAGE Encyclopedia of Research Design, SAGE Publications. https://doi.org/10.4135/9781412961288.n228
51. Harzing AW (2006) Response Styles in Cross-national Survey Research: A 26-country Study, Int J Cross Cult Manag, 6(2), 243–266. https://doi.org/10.1177/1470595806066332
52. Kemmelmeier M (2016) Cultural differences in survey responding: Issues and insights in the study of response biases, Int J Psychol, 51(6), 439–444. https://doi.org/10.1002/ijop.12386
53. Ulitzsch E, Henninger M, Meiser T (2024) Differences in response-scale usage are ubiquitous in cross-country comparisons and a potential driver of elusive relationships, Sci Rep, 14, 10890. https://doi.org/10.1038/s41598-024-60465-0