UDC:159.9.072:711.4(599)
doi: 10.5379/urbani-izziv-en-2025-36-02-05

Received: 31 August 2025
Accepted: 18 November 2025

Zaldy F. CORPUZ

Development and initial validation of the Urban Restorative Potential Scale: Evidence from the Philippines

Urban environments influence mental well-being, yet most restoration scales have been developed for natural settings. This study developed and initially validated the Urban Restorative Potential Scale (URPS), a design-oriented measure for everyday urban contexts. A set of twenty-five items grounded in attention restoration theory and stress reduction theory was tested with 1,001 residents of Urdaneta, Philippines. Confirmatory factor analysis supported a seventeen-item, four-factor structure (quality, functionality, captivation, and relaxation) under a second-order urban restorative potential (URP) factor, with good fit and reliability. Convergent validity was adequate; subscale correlations were high but consistent with a meaningful higher-order construct. The URP factor strongly predicted overall satisfaction with the scene viewed. Multi-group models indicated context dependence across urban scene types (USTs): quality loaded most strongly on URP in recreational scenes; captivation dominated in housing, institutional, streetscape, and commercial scenes; functionality was most influential in transit scenes; and relaxation was an important but not leading factor for all USTs. The findings are cross-sectional; therefore, causal claims are not warranted. The URPS offers a practical diagnostic profile for design hypotheses, highlighting that what "feels restorative" varies by scene type.

Keywords: restorative environments, urban scene types, psychometric validation, measurement invariance, environmental psychology

1 Introduction

Urban environments often elevate stress, anxiety, and attention fatigue due to noise, crowding, and limited access to psychologically supportive spaces (Bowler et al., 2010; McDonald et al., 2018). However, most tools for assessing "restorativeness" were initially developed for parks and forests, not for streets, housing blocks, or transit stops (Hartig et al., 1997). This gap creates a demand for psychologically supportive everyday environments, not merely green infrastructure. This raises a critical design question: can ordinary urban places, such as plazas, streets, residential blocks, campuses, commercial districts, and transit hubs, offer mental relief?

Traditional restorative environment research highlights settings that support recovery from attention fatigue and stress (Kaplan, 1995; Ulrich et al., 1991). Attention restoration theory (ART) proposes that recovery occurs when environments provide four key qualities: being away (psychological distance from routine demands), fascination (effortless attention, particularly "soft fascination", which gently engages without mental strain), extent (sufficient scope and coherence to feel immersed), and compatibility (support for intended activities and purposes; Kaplan & Kaplan, 1989; Kaplan, 1995). Stress reduction theory (SRT) emphasizes rapid physiological recovery in environments that feel safe, legible, and aesthetically pleasant (Ulrich, 1984; Ulrich et al., 1991).

Although both theories were originally developed in relation to natural settings, subsequent studies have demonstrated that certain urban environments can also foster restoration (San Juan et al., 2017; Lindal & Hartig, 2013; Bornioli & Subiza-Pérez, 2023). Features such as visual interest, legibility, perceived safety, and opportunities to pause or find refuge have been shown to support attentional relief and stress recovery in everyday urban contexts (Peschardt & Stigsdotter, 2013; Lindal & Hartig, 2015; Joye & van den Berg, 2011). In short, restoration is not confined to forests or wilderness; it may also emerge in ordinary urban niches, such as the shaded corner of a transit interchange, when those places feel ordered, comprehensible, and welcoming.

The long-standing "natural versus urban" dichotomy in restoration research is increasingly recognized as a gradient rather than a binary distinction. Although meta-analytic evidence still finds natural settings more restorative on average, the wide variability within and across settings cautions against simple either-or classifications (Menardo et al., 2021). The Perceived Restorativeness Scale (PRS), which operationalizes the four components of ART (being away, fascination, extent, and compatibility), has been the dominant measurement tool since its development and validation in natural settings (Hartig et al., 1997; Pasini et al., 2014). However, the PRS was primarily designed to assess experiences of immersion in nature or escape from urban demands. In everyday urban contexts, many PRS items presume a "retreat" experience (e.g., being away interpreted as physical escape), whereas much real-world urban restoration involves micro-restoration, which is brief moments of relief occurring on-site while navigating daily activities (Joye & Dewitte, 2018).

More fundamentally, the PRS underspecifies design-relevant features that shape urban restoration. Although ART's qualities of fascination and being away capture important psychological mechanisms, they do not directly address the physical environmental qualities that urban users consistently identify as critical to psychological comfort. Research on urban environmental preferences reveals that people value specific tangible features such as cleanliness and maintenance (visual order), functional amenities such as seating and shade (physical comfort), safety and legibility (navigability), and aesthetic appeal (visual engagement) as determinants of whether a space feels supportive or stressful (Nordh et al., 2009; Peschardt & Stigsdotter, 2013; Ríos-Rodríguez et al., 2021; Jennings & Bamkole, 2019). These features align conceptually with ART's extent and compatibility but require more explicit operationalization for urban design contexts. Similarly, SRT's emphasis on safety and positive affect points to the importance of perceived care and environmental quality, aspects largely implicit in nature-based scales. The absence of items directly capturing these built-environment attributes is a significant gap: designers cannot easily translate abstract concepts such as being away or fascination into specific interventions, but they can modify maintenance regimes, add seating, improve wayfinding, or enhance visual interest. Thus, an Urban Restorative Potential Scale (URPS) must bridge the gap between psychological theory and observable, designable features.

This highlights two persistent gaps. First, there is apparently no validated scale that explicitly operationalizes restorative potential using urban design terminology accessible to practitioners yet. Second, few studies have tested whether restorative processes operate equivalently across diverse urban settings, leaving open the question of whether restoration is a universal phenomenon or context dependent. Addressing these gaps requires moving beyond adaptation of nature-centric measures toward purpose-built instruments that treat urban environments as legitimate objects of restoration research.

Urban restoration likely varies by context. Different urban scene types (USTs), such as residential streets, institutional campuses, commercial districts, streetscapes, transit hubs, and recreational parks, exhibit distinct affordances and stressors that may shape restorative profiles (Karmanov & Hamel, 2008; Wilkie & Clements, 2018). Evidence shows that greenery supports restoration in residential streets (Lindal & Hartig, 2013, 2015; Zhao et al., 2020), institutional courtyards (Cooper Marcus & Sachs, 2014), pedestrian squares (Subiza-Pérez et al., 2020), and parks (Korpela et al., 2010), whereas transit environments often elevate stress unless specific design mitigations (e.g., clear wayfinding, seating, and shelter) are introduced (Evans & Wener, 2007). This heterogeneity raises a methodological question: few studies have tested measurement invariance across scene types, leaving uncertain whether restoration operates through similar mechanisms universally or whether the relative importance of different restorative dimensions shifts with context (Chen, 2007; Milfont & Fischer, 2010; Bornioli & Subiza-Pérez, 2023).

To address these gaps, the URPS was developed, operationalizing ART and SRT through four design-oriented factors. Quality encompasses environmental care, order, and safety (SRT's non-threatening environments, ART's extent). Functionality reflects physical affordances (seating, shade, and wayfinding) that support intended activities (ART's compatibility). Captivation embodies soft fascination through visual interest and architectural engagement (ART's fascination). Relaxation reframes being away as psychological refuge achievable on-site, encompassing calmness within urban settings (ART's being away, SRT's stress reduction). Thus, URPS retains ART's attentional mechanisms while operationalizing built environment attributes underspecified in nature-based scales. URPS was tested across six USTs (recreational, housing, institutional, streetscape, commercial, and transit), given evidence that restorative potential varies by context: wayfinding matters in transit hubs, visual richness in streetscapes, and order in parks (Evans & Wener, 2007; Peschardt & Stigsdotter, 2013; Purcell et al., 2001).

This study pursues three aims: 1) to test the URPS factor structure, model fit, reliability, and selected aspects of validity (convergent, discriminant, and criterion); 2) to examine measurement invariance across USTs using multi-group confirmatory factor analysis, testing whether the URPS operates equivalently or context dependently; and 3) to assess criterion validity by testing whether URPS scores predict overall satisfaction with the environment.

2 Methodology

2.1 Research design and participants

A cross-sectional, between-subjects image-rating study was conducted in Urdaneta, Philippines, in June 2025. Each participant viewed one randomly assigned urban scene photo and rated each scene. This between-subjects design avoids carryover and contrast effects common in within-subjects image ratings (Poulton, 1982; Aguinis & Bradley, 2014), which produces a two-way dataset (person × one scene). It also allows for each participant to provide a focused, unbiased assessment of a single environment. Image-based assessment is appropriate for scale development because it permits systematic control of visual features while maintaining ecological validity (Stamps, 2010).

A total of 1,070 city residents participated, recruited through on-site convenience sampling with quota targets for age groups to ensure a broad demographic mix. Inclusion criteria required that participants currently lived in Urdaneta and could give informed consent. The study excluded individuals with uncorrected visual impairment, those unable to follow task instructions, and those that were visibly intoxicated. After data quality screening, the final analytic sample was comprised of 1,001 participants. The sample was 58.3% female and 40.9% male (0.8% did not specify sex). Age was distributed across four life stages: adolescents and emerging adults (16–19 years, 29.6%), young adults (20–34, 31.7%), middle-aged adults (35–54, 29.1%), and older adults (55+, 9.6%). These age bands reflect developmental stages relevant to environmental perception (Pan American Health Organization, 2020) and ensured balanced representation across the lifespan. Education levels were distributed to no formal schooling (1%), elementary school (10.1%), high school (39.5%), bachelor's degrees (40.6%) and graduate degrees (8.8%). Participants came from rural areas (41.1%), small towns (19.7%), suburban areas (20.6%), and urban centres (18.6%). They were randomly assigned to view one image showing one of the six USTs, with their shares distributed relatively evenly across the UST categories (14.9% to 18.2% of participants per UST).

2.2 Stimuli and procedure

The visual stimuli were drawn from an extensive photo bank assembled specifically for this project. Thousands of candidate photos of urban scenes were taken by a research team during multiple field visits. To maximize generalizability, images were taken across a rural–urban gradient. Figure 1 shows the locations where they were taken. Approximately two-thirds (66.1%) of the photos came from sites within Pangasinan, 17.9% from Metro Manila, 14.2% from the nearby province of Nueva Ecija, and the remaining 1.8% from Tarlac. All photos were taken under comparable conditions to control for lighting and sky conditions. The photos were taken using standard focal lengths (24–50 mm full-frame equivalent) with minimal post-processing (horizon levelling and exposure normalization only) to preserve a natural appearance. Display settings were

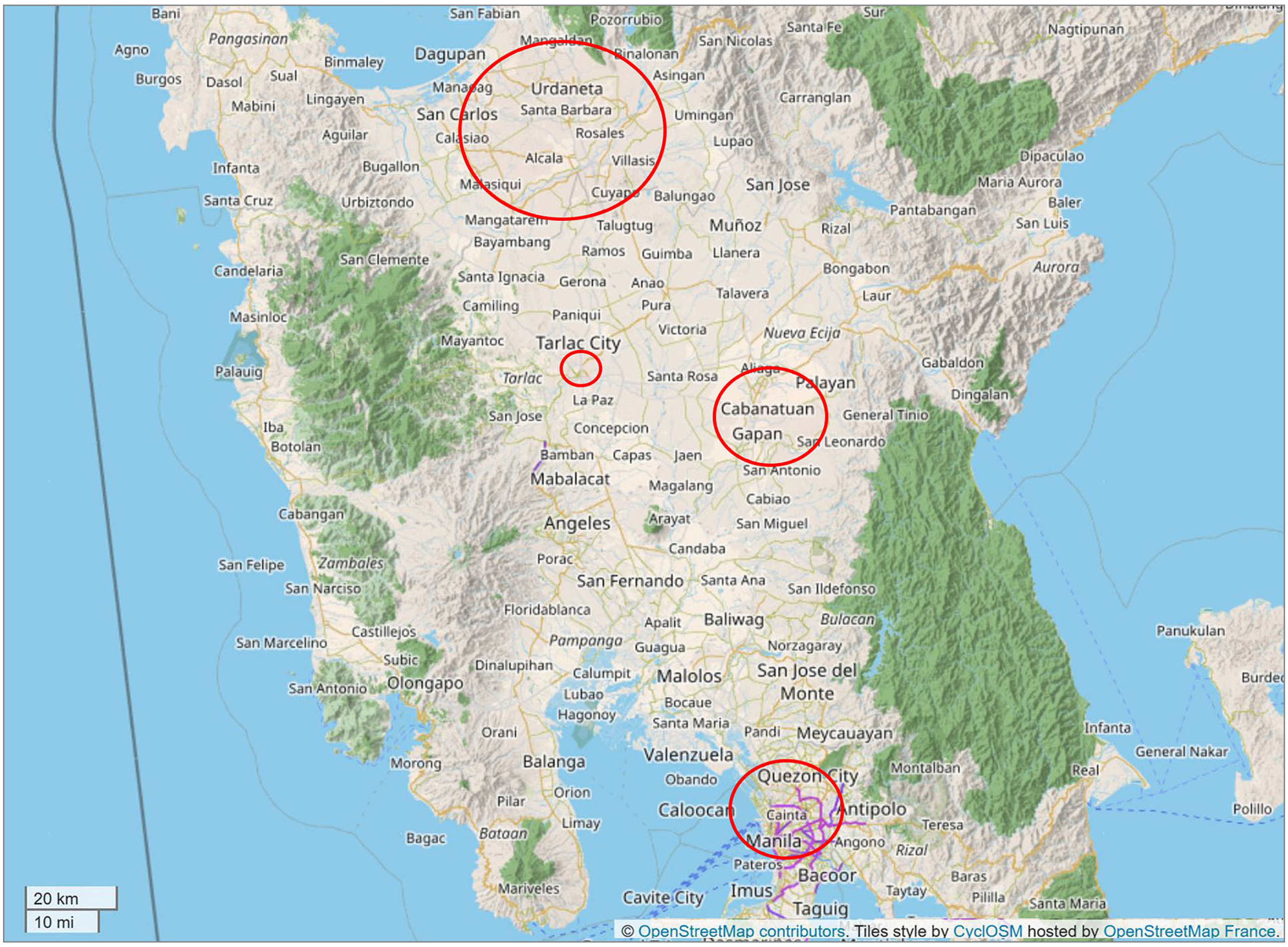

Figure 1: Locations of where the photos were taken and used as stimuli for the study (base map: OpenStreetMap).

standardized across sessions (auto-brightness disabled, screen brightness ≥ 50%).

From this pool, a balanced set of images was selected representing the six USTs of interest: recreational green spaces, residential areas, institutional/campus settings, streetscapes, commercial/business districts, and transit-oriented places. For each urban scene type, nineteen representative images (114 total) were selected through a systematic rating process. Six trained urban design students independently evaluated candidate photos for visible restorativeness indicators (e.g., cleanliness, greenery, seating, and visual order) versus detractors (e.g., litter, clutter, and lack of refuge). This process ensured variability within each UST, including both higher- and lower-restorative exemplars rather than only idealized scenes. Prior to the main study, it was validated that the participants would perceive the images as intended. Thirty independent raters (urban design students) sorted a sample of the photos into UST categories without location information. Inter-rater reliability was substantial (Fleiss's kappa = 0.72, 95% CI [0.68, 0.76]), confirming clear scene type recognition. Figure 2 shows examples of the stimuli, one for each category.

Data collection took place at community halls (barangays) across Urdaneta. To ensure standardized conditions, facilitators completed a pre-session checklist verifying a quiet environment (< 30 dBA ambient noise), glare-free display, appropriate viewing distance (50–70 cm), and adherence to the standardized script.

Upon arrival, participants received information about the study and provided informed consent. Each participant was then randomly assigned one photo from the pool of 114 images, in which participants were relatively evenly distributed across the six USTs. Participants viewed their assigned scene on a laptop screen for sixty seconds while facilitators ensured attention to the image. Immediately following image exposure, participants supplied their demographic information, including age group, sex, highest educational attainment, and current residence location, and then they rated the URPS and overall satisfaction items either on a tablet or on paper, depending on their preference and literacy level. Trained data collectors monitored all sessions to ensure comprehension and protocol adherence.

Figure 2: Example study stimulus per urban scene type: a) recreational scene; b) housing scene; c) institutional scene; d) streetscape scene; e) commercial scene; f) transit scene (photos: author).

The protocol was reviewed and approved by the University Research Ethics Committee in accordance with the National Ethics Guidelines for Health and Social Research (Philippine Health Research Ethics Board, 2022). Written informed consent was obtained from all participants. For participants under eighteen, guardian consent and participant assent were obtained in accordance with ethical guidelines for research involving minors. Responses were anonymized and processed under the Data Privacy Act of 2012 (RA 10173).

2.3 Measures

The URPS was developed as a self-report instrument to assess the perceived restorativeness of urban environments. A theory-driven approach to item generation was adopted, drawing directly on principles from ART and SRT to ensure content validity. Rather than adapting items from existing nature-based scales, twenty-five original items tailored to urban design features were developed while preserving core restorative principles. Each item presented a statement evaluating attributes of the scene's restorative quality, rated on a five-point Likert scale (1 = strongly disagree to 5 = strongly agree).

The twenty-five items were organized a priori into four conceptual factors based on the theoretical framework. Quality encompasses aesthetic upkeep and environmental coherence that contribute to perceived safety and order (example item: "The environment appears clean and well maintained"). Functionality covers physical amenities, comfort features, and spatial affordances that support user activities (example item: "This environment provides suitable places for sitting and resting"). Captivation reflects the scene's ability to attract and hold attention effortlessly, linked to the concepts of fascination and visual complexity (example item: "The visual details in this environment are noticeable and appealing"). Relaxation encompasses one's sense of calmness, refuge, and psychological distance, corresponding to the being away dimension of restorative experience (example item: "This environment feels peaceful and calm to me").

Prior to full deployment, a pilot study with fifty university students and staff tested the clarity and comprehensiveness of all twenty-five items. Participants rated eighteen trial images using the draft questionnaire and provided feedback via a debrief form with clarity (five-point) ratings and open comments. Based on this feedback, technical or unusual terms were revised with local translations added in parentheses to improve comprehension, particularly for participants with limited education or English proficiency. All twenty-five items were retained after these linguistic refinements, and the administration script and questionnaire were finalized for the main study. No items were reverse coded.

A single-item criterion measure assessed overall environmental satisfaction: "Overall, I am satisfied with the quality of this environment." Participants rated this item on the same five-point Likert scale. This item served as an external criterion to test whether URPS scores predict general environmental appraisal.

2.4 Data collection and screening

Responses were recorded via Google Forms; paper forms were used when preferred and double entered for accuracy. Google Forms enforced required responses and value validation (Likert items restricted to 1–5). Paper forms were checked at point-of-collection and double entered with an independent second pass; discrepancies were reconciled. Data were exported to MS Excel and screened for quality. The combination of required digital form fields and double-entry verification for paper forms eliminated missing data. Data quality screening focused on identifying inattentive responding patterns. The study screened for inattentive responding (i.e., low-effort or "straight-lining" patterns) by calculating within-person response variability across all URPS items rated for the single scene viewed. Participants whose URPS responses showed extremely low variability (standard deviation ≤ 0.40 on a five-point scale) were excluded as likely inattentive. This threshold corresponds to responses varying by less than half a scale point on average and represents near-uniform responding (e.g., marking all 3s or 4s with minimal variation), consistent with recommended quality-control practices (Huang et al., 2012). This removed approximately 6.5% of the initial cases, yielding a final analytical sample of 1,001.

2.5 Statistical analysis

Analyses were conducted in two phases using IBM SPSS Amos (version 21.0.0). A significance level of $\alpha = 0.05$ was adopted as the criterion for hypothesis testing.

The first phase specified a theory-driven four-factor model (twenty-five items) and estimated it using confirmatory factor analysis (CFA). It used maximum likelihood (ML) estimation, treating five-point Likert items as approximately continuous, consistent with simulation evidence (Rhemtulla et al., 2012; see also Li, 2016; Norman, 2010). Given the sample size and the distribution of responses, ML estimation performed reliably. Model fit was assessed using common criteria ($\chi^2/df \leq 3$, CFI ≥ 0.95, SRMR ≤ 0.08, RMSEA ≤ 0.06–0.08; Gaskin & Lim, 2016; Hu & Bentler, 1999). Modification indices were reviewed to identify items with weak loadings (< 0.50; Hair et al., 2010; Kline, 2016; Byrne, 2016) or redundancy, and items were removed only when the change was theoretically supported. This was followed by a comparison of alternative model structures using overall fit indices (CFI, SRMR, and RMSEA) and parsimony-oriented information criteria (BIC and sample-size-adjusted BIC [SABIC]). Four possible structures were compared: a single-factor model, a correlated four-factor model (quality, functionality, captivation, and relaxation), a second-order model in which the four factors load onto one higher-order factor, and a bifactor model with one general factor plus the four specific factors. The model was judged using BIC and SABIC, in which lower values indicate a better balance of fit and simplicity (Burnham & Anderson, 2002). AIC was also checked and showed the same ranking. For the selected model, reliability (α and composite reliability ≥ 0.70), convergent validity (AVE ≥ 0.50; Fornell & Larcker, 1981), and discriminant validity using HTMT (< 0.85; Henseler et al., 2015) with percentile-bootstrap 95% confidence intervals for the HTMT ratios were examined.

The second phase examined measurement invariance across the six USTs using sequential multi-group CFA of the final structure. It tested configural invariance (same factor structure across groups), followed by metric invariance (equal factor loadings) and structural invariance (equal factor correlations/variances). Scalar invariance was not tested because the focus was on whether the measurement model structure replicates across contexts rather than comparing latent means. Given that chi-square difference tests are sensitive to large samples (Yuan & Bentler, 2004), practical fit changes were emphasized: ΔCFI < 0.01 and ΔRMSEA < 0.015 indicated acceptable invariance (Chen, 2007; Cheung & Rensvold, 2002). When only a subset of parameters could be constrained equally while maintaining acceptable fit, partial invariance was documented and interpretations were adjusted accordingly (Byrne, 2016; Putnick

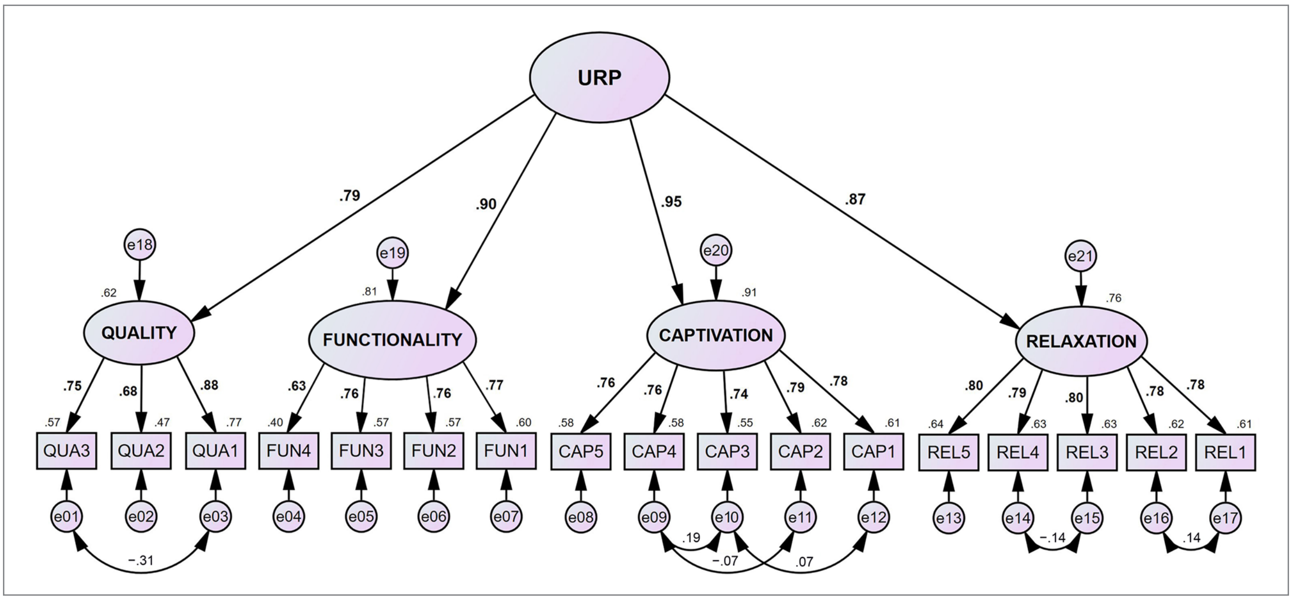

Figure 3: The retained and most optimal second-order URP measurement model solution with standardized estimates (diagram generated in IBM SPSS Amos).

Table 1: Fit indices of rival measurement models.

Model	df	χ^2	χ^2/df	CFI	SRMR	RMSEA	BIC	SABIC
Correlated four-factor	113	356.90	3.158	0.975	0.030	0.046	633.25	506.21
Second-order (URP)	109	307.73	2.823	0.980	0.027	0.043	611.714	471.97
Bifactor	115	358.60	3.119	0.975	0.030	0.046	631.68	500.44
Unidimensional	119	1,309.26	11.002	0.878	0.058	0.100	1,544.16	1,436.17

Note: df = degrees of freedom, χ^2 = chi-squared statistic, χ^2/df = ratio of chi-squared to degrees of freedom, CFI = comparative fit index, SRMR = standard root mean square residual, RMSEA = root mean square error of approximation, BIC = Bayesian information criterion, SABIC = Schwarz's adjusted Bayesian information criterion.

Table 2: Reliability and convergent validity statistics for the URPS subscales.

Factor	α	CR	AVE	MSV	MaxR(H)
Quality	0.793	0.792	0.561	0.575	0.810
Functionality	0.818	0.821	0.535	0.722	0.879
Captivation	0.878	0.878	0.590	0.722	0.829
Relaxation	0.893	0.893	0.626	0.653	0.894

Note: α = coefficient alpha, CR = composite reliability, AVE = average variance extracted, MSV = maximum shared variance, MaxR(H) = maximum reliability (H). Adequate reliability: α and CR ≥ 0.70, convergent validity: AVE ≥ 0.50, discriminant validity: AVE > MSV.

Table 3: Discriminant validity: heterotrait–monotrait (HTMT) ratios.

	Quality	Functionality	Captivation	Relaxation
Quality	—			
Functionality	0.730	—		
Captivation	0.738	0.836	—	
Relaxation	0.687	0.771	0.807	—

Note: HTMT < 0.85 suggests adequate discriminant validity. All values meet this criterion, indicating that the four factors are empirically distinguishable despite their theoretical interrelatedness within the higher-order URP construct.

& Bornstein, 2016). To characterize context-specific patterns, first- and second-order factor loadings were estimated separately for each UST. This allowed examination of which factors and items show stronger or weaker relationships to overall restorativeness across different urban contexts.

3 Results

3.1 Model refinement and fit

The initial twenty-five–item CFA showed acceptable model fit (χ^2 = 1014.114, df = 241, χ^2/df = 4.208, CFI = 0.951, SRMR = 0.034; RMSEA = 0.057). Based on modification indices and theoretical considerations, eight items were removed due to weak factor loadings (< 0.50), high cross-loadings, or content redundancy. For example, multiple items addressing visual variety were consolidated to eliminate overlap, and items with double-barrelled phrasing were deleted to improve clarity. This yielded a seventeen-item solution comprising three quality items, four functionality items, five captivation items, and five relaxation items. All retained indicators loaded strongly on their intended latent factors (0.630–0.878; all $p < 0.001$). The four first-order factors were substantially intercorrelated, supporting the presence of a higher-order construct representing the URP (Figure 3). The RMSEA close-fit test was non-significant (PClose = 0.984), indicating close fit (RMSEA ≤ 0.05).

The study compared four rival measurement models (Table 1). The unidimensional model showed poor fit, confirming that URP cannot be treated as a single undifferentiated construct. Three multidimensional alternatives (a correlated four-factor model, a bifactor model, and a second-order hierarchical model) all demonstrated good fit. Information criteria favoured different models: AIC was lowest for the bifactor model, whereas BIC and SABIC were lowest for the second-order model. The second-order model was retained as the final structure based on three considerations: a) marginally superior absolute fit indices (CFI, SRMR, and RMSEA), b) stronger performance on parsimony-adjusted criteria (BIC and SABIC), and c) theoretical interpretability as a general URP construct manifested through four conceptually distinct factors.

3.2 Reliability and validity

All four URPS subscales showed strong internal consistency (Table 2), with coefficient alpha and composite reliability ranging from 0.793 to 0.893. Convergent validity was supported: AVE exceeded 0.50 for all factors, meaning that each latent factor explains more than half of the variance in its indicators (Fornell & Larcker, 1981).

Discriminant validity evidence supported the factor structure while also justifying the second-order model (Table 3). All HTMT ratios fell below 0.85 (range: 0.687–0.836), indicating adequate discriminability among the four factors. However, the Fornell–Larcker criterion (AVE > MSV) was not satisfied for all factors, indicating substantial shared variance among subscales. This pattern is theoretically consistent with a higher-order factor structure: the factors are conceptually distinct yet strongly intercorrelated because they all reflect aspects of the broader URP construct.

Table 4 presents the final seventeen-item URPS, organized by factor, with item codes and shortened descriptions used in analyses and subsequent reporting.

3.2.1 Criterion validity

To assess whether URPS scores predict overall environmental satisfaction, the satisfaction criterion item was regressed on the second-order URP factor. The relationship was strong and significant (β = 0.888, $p < 0.001$, $R^2 = 0.788$), indicating that URP accounts for approximately 79% of variance in overall satisfaction. This provides evidence that the URPS captures aspects of environmental quality that meaningfully predict global environmental appraisal.

3.2.2 Common method bias

Given that all variables were self-reported in a single session, potential common method bias was assessed using two approaches. Harman's single-factor test revealed that the first unrotated factor accounted for only 32% of variance, well below the 50% threshold indicating problematic bias. In addition, adding an unmeasured latent method factor to the CFA did not improve model fit or substantively alter factor loadings, suggesting that common method variance did not drive the observed factor structure.

3.3 Measurement invariance test

Measurement invariance was tested to determine whether the URPS structure and factor loadings operate equivalently across the six USTs. Multi-group CFA results are presented in Table 5.

The lack of metric and structural invariance indicates that the URPS operates differently across urban contexts. Specifically, factor loadings and the relative importance of the four factors in defining overall URP vary significantly by scene type. This context dependency is illustrated in Tables 6 and 7, which show how different restorative factors dominate in different urban settings.

Table 4: Final seventeen-item URPS with factor assignment and item wording.

Factor	Code	Short description	Questionnaire item
Quality	QUA1	Appears clean and maintained	The environment appears clean and well-maintained.
	QUA2	Layout is logical and orderly	The arrangement and layout of this space seem logical and orderly.
	QUA3	Visually appealing	I find this environment visually appealing.
Functionality	FUN1	Provides places to sit and rest	This environment provides suitable places for sitting and resting.
	FUN2	Amenities are functional and well kept	Amenities such as benches, signs, and lighting fixtures look well kept and functional.
	FUN3	Offers appealing leisure activities	There are appealing leisure opportunities available here.
	FUN4	Provides shelter or shade	This environment offers sufficient shelter or shade.
Captivation	CAP1	Features diverse spatial elements	This environment has diverse spatial features.
	CAP2	Visually rich with distinct elements	There is a high level of visual interest due to many distinct elements in this scene.
	CAP3	Appealing visual details stand out	The visual details in this environment are noticeable and appealing.
	CAP4	Unique details hold my attention	Unique elements or surprising details in this environment hold my attention easily.
	CAP5	Colour variety is stimulating	The colour variety in this environment is visually stimulating.
Relaxation	REL1	Feels peaceful and calm	This environment feels peaceful and calm to me.
	REL2	Greenery creates a relaxing atmosphere	The greenery or natural elements I see here create a relaxing atmosphere.
	REL3	Openness and views feel calming	The openness and views in this environment make me feel at ease.
	REL4	Lighting and colours are soothing	The lighting and colour scheme in this environment soothe my eyes and mind.
	REL5	Has comfortable retreat spaces	There are spaces here where I would feel comfortable retreating and relaxing.

Note: All items were rated on five-point Likert scale (1 = strongly disagree, 5 = strongly agree).

Table 5: Nested model comparison (Δ vs. previous) for invariance across urban scene types.

Model	Δdf	$\Delta\chi^2$	p	ΔCFI vs. prev	ΔRMSEA vs. prev	Interpretation
Configural (baseline)	—	—	—	—	—	Acceptable baseline fit
Metric (equal loadings)	65	149.436	< 0.001	–0.011	+0.001	Not supported
Structural (equal covariances)	15	35.375	< 0.002	–0.020	+0.002	Not supported

Note: Comparison statistics (Δ) are calculated relative to the less constrained model at each step. Decision criteria: ΔCFI ≤ 0.010 and ΔRMSEA ≤ 0.015 indicate acceptable invariance. Configural invariance was supported (CFI = 0.934, RMSEA = 0.030), indicating that the four-factor structure fit adequately across all six USTs. However, metric invariance was not supported: constraining factor loadings to equality across USTs resulted in significant fit deterioration ($\Delta\chi^2$ (65) = 149.436, $p < 0.001$, ΔCFI = –0.011). Similarly, structural invariance was not achieved ($\Delta\chi^2$ (15) = 35.375, $p = 0.002$, ΔCFI = –0.020).

Table 6: Factor importance varies by urban scene type: standardized loadings on second-order URP factor.

Urban scene type	Quality (β)	Functionality (β)	Captivation (β)	Relaxation (β)
Recreational	**0.908**	0.858	0.815	0.872
Housing	0.847	0.867	**0.979**	0.901
Institutional	0.819	0.761	**0.875**	0.810
Streetscape	0.841	0.914	**0.954**	0.764
Commercial	0.783	0.972	**0.999**	0.912
Transit	0.683	**0.971**	0.908	0.881

Note: Values are standardized regression weights of first-order factors on the second-order URP factor. Bold values indicate the most salient factor for each scene type.

Table 7: Standardized loadings (β) of items on first-order factors across urban scene types.

Item (shortened)	Urban scene types					
	Recreational	Housing	Institutional	Streetscape	Commercial	Transit
Quality						
Appears clean and maintained	0.724	0.576	0.726	0.731	0.693	0.760
Visually appealing	0.702	**0.834**	**0.852**	**0.811**	**0.868**	**0.836**
Layout is logical and orderly	**0.730**	0.627	0.682	0.799	0.576	0.754
Functionality						
Amenities are functional and well kept	0.612	**0.786**	0.757	0.715	**0.804**	0.699
Provides places to sit and rest	**0.888**	0.732	**0.772**	0.690	0.760	0.760
Offers appealing leisure activities	0.667	0.773	0.620	**0.838**	0.752	0.711
Provides shelter or shade	0.599	0.661	0.603	0.632	0.623	**0.764**
Captivation						
Appealing visual details stand out	0.663	0.775	0.793	0.757	**0.811**	**0.799**
Features diverse spatial elements	0.615	0.752	**0.821**	0.792	0.719	0.798
Visually rich with distinct elements	0.692	0.778	0.744	**0.817**	0.807	0.760
Unique details hold my attention	0.679	0.734	0.793	0.653	0.720	0.760
Colour variety is stimulating	**0.744**	**0.797**	0.735	0.782	0.798	0.798
Relaxation						
Feels peaceful and calm	0.773	0.706	0.762	0.751	0.831	0.788
Greenery creates a relaxing atmosphere	**0.741**	**0.823**	0.854	0.704	0.799	0.703
Openness and views feel calming	0.730	0.800	**0.864**	**0.802**	0.739	0.743
Has comfortable retreat spaces	0.686	0.744	0.705	0.794	**0.845**	**0.880**
Lighting and colours are soothing	0.740	0.738	0.763	0.791	0.829	0.705

Note: Values are standardized regression weights of items on their designated first-order factors.

Table 7 presents item-level loadings by scene type (bolded values indicate the highest within each factor). For example, visually appealing is the strongest quality indicator except in recreational settings, where logical layout dominates.

In summary, whereas the URPS demonstrated strong psychometric properties overall, the lack of measurement invariance across urban scene types indicates that restorative potential manifests differently in different contexts. This finding has important implications for how urban environments should be assessed and designed, as discussed in Section 4.

4 Discussion

Urban environments can support psychological restoration, but existing assessment tools were designed primarily for natural settings. This study developed and initially validated the URPS, a theory-driven instrument tailored to urban design features. Drawing on ART and SRT, it identified four factors of URP: quality (aesthetic care and visual order), functionality (physical amenities and comfort), captivation (visual interest and soft fascination), and relaxation (calmness and psychological refuge). CFA supported this structure, revealing that the four factors are conceptually distinct yet empirically interrelated, a pattern consistent with the notion that restorative urban environments engage multiple psychological pathways simultaneously. These four factors align with established environmental psychology constructs while capturing urban-specific features often overlooked in nature-based scales.

Psychometric analyses demonstrated strong reliability and validity evidence. All four subscales showed high internal consistency, adequate convergent validity, and discriminability. The factors were substantially intercorrelated, supporting a second-order hierarchical model wherein quality, functionality, captivation, and relaxation represent distinct but related facets of a broader urban restorative potential (URP) construct. This hierarchical structure has practical utility: the total URPS score provides an overall assessment of restorative potential, and subscale profiles reveal specific strengths and weaknesses (e.g., "high functionality but low captivation"). The strong correlations among factors suggest that highly restorative urban environments typically excel across multiple factors rather than compensating for deficiencies in one area by excelling in another.

The URPS demonstrated strong criterion validity: the second-order URP factor predicted 79% of variance in overall environmental satisfaction. This substantial relationship indicates that the URPS captures perceptual qualities central to how people evaluate urban environments. Although this cross-sectional association does not establish causality, it provides partial evidence that URPS scores reflect meaningful and preference-relevant aspects of environmental quality.

4.1 The context-dependent nature of urban restorative perceptions

Restorative potential manifests differently across urban contexts. Measurement invariance testing revealed that respondents weighted the four URPS factors differently by scene type. In recreational settings (parks and green spaces), quality (reflecting cleanliness, order, and aesthetic care) was paramount, likely because these spaces serve contemplative and leisure functions, where environmental maintenance signals safety and care. Captivation dominated in housing, institutional, streetscape, and commercial contexts, where visual interest and architectural engagement may compensate for limited natural elements. In transit environments, functionality was most critical, consistent with research showing that basic comfort and wayfinding reduce transit-related stress (Evans & Wener, 2007). Relaxation was valued across all settings, but it never emerged as the sole dominant factor, suggesting it operates as a necessary but insufficient condition for restorative potential.

These findings challenge the notion of a universal formula for restorative urban design. Instead, effective interventions must be tailored to context: enhancing visual order in parks, introducing architectural interest in streetscapes, or improving wayfinding and seating in transit hubs. Although the cross-sectional design in this study limits causal inference, these patterns provide evidence-informed guidance for prioritizing design interventions in different urban settings.

4.2 Theoretical and practical implications

The findings extend restoration theory in three ways. First, the URPS shows that core restorative factors can be meaningfully represented in people's appraisals of everyday urban environments, not only natural or park settings. The initially validated four-factor structure shows that urban restorativeness involves both ART mechanisms (captivation as soft fascination and relaxation as being away) and built environment qualities absent from nature-based scales (quality as aesthetic care and functionality as physical affordances). Second, the lack of measurement invariance across urban scene types suggests that perceived restorative potential is context dependent rather than universal. This aligns with recent calls to move beyond nature-centric restoration models (Joye & Dewitte, 2018; Bornioli & Subiza-Pérez, 2023) and supports the notion of multiple pathways to restoration (Menardo et al., 2024). Different environments restore through different mechanisms: parks through maintained order and greenery, transit hubs through legibility and comfort, and commercial streets through visual engagement. Third, the hierarchical factor structure indicates that, within these data, scenes judged as highly restorative tend to score well across multiple attributes rather than a single feature. The strong intercorrelations among factors indicate that highly restorative environments typically excel across factors, challenging single-mechanism explanations (e.g., "just add greenery") common in popular discourse.

The URPS offers a diagnostic tool for urban design practice, identifying specific intervention areas. Rather than applying generic "restorative design" principles uniformly, practitioners can use URPS profiles to target context-specific improvements: 1) recreational settings could perhaps prioritize maintenance, visual order, and legibility (quality); 2) residential, institutional, and commercial areas could enhance architectural interest, art, and visual variety (captivation); 3) transit environments could focus on wayfinding, seating, shelter, and basic comfort (functionality); and 4) in all settings, there is a need to ensure opportunities for psychological refuge and calm (relaxation as baseline).

Importantly, the findings suggest that urban environments need not mimic natural settings to be experienced as restorative. Even high-traffic commercial streets or transit stops can provide psychological relief when designed with attention to context-appropriate features. For example, a transit hub with clear wayfinding, comfortable seating, and adequate shade or shelter may be more restorative than a poorly maintained park.

Because the URPS factors showed non-invariance, caution is advised against using total scores to rank or compare different urban scene types (e.g., "parks score higher than transit stops"). Instead, the scale is best applied within scene types to compare alternative designs or track changes over time. The URPS provides a framework for evidence-informed design decisions while recognizing that restorativeness is not a universal formula, but a context-dependent attribute shaped by setting-specific affordances and user needs.

4.3 Strengths, limitations, and future research

4.3.1 Strengths

This study has several notable strengths. First, the URPS addresses a recognized gap by providing an initially validated restorativeness scale explicitly designed for urban environments rather than adapted from nature-based measures. Second, the large sample and diverse age range provide stable parameter estimates and broad representation. Third, comprehensive psychometric testing, including rival model comparisons, validity checks, and measurement invariance tests, provided strong evidence for the scale's structure and properties. Fourth, the balanced stimulus set (nineteen images × six scene types) and between-subjects design eliminated carryover effects while allowing robust multi-group comparisons. Finally, standardized viewing conditions and attention checks enhanced data quality and internal validity.

4.3.2 Limitations

Several limitations warrant consideration. Regarding sampling and generalizability, participants were recruited via convenience sampling in a single Philippine city, limiting population generalizability. Cross-cultural replication is needed to determine whether the factor structure and context-specific patterns replicate in other cultural contexts. Regarding the measurement approach, the stimuli were static photos, which capture visual features but omit multisensory, temporal, and social dimensions of lived experience. Although this approach maximizes experimental control, field-based validation is needed to confirm whether URPS scores predict restoration outcomes in situ. In addition, single-session self-report data limit conclusions about test–retest reliability and introduce potential for common method bias, although the statistical checks found minimal evidence of such bias.

Regarding scale development, although the four URPS factors are theory grounded and demonstrated good psychometric properties, the scale may not capture all dimensions of urban restorativeness. Some captivation items may read as stronger stimulation in certain contexts; future refinements could soften this language while retaining the validated structure. The lack of measurement invariance across scene types, while theoretically meaningful, means that cross-context mean comparisons are not psychometrically warranted. Finally, this study provides initial evidence of reliability and selected aspects of validity, but external validation against independent restoration measures (e.g., physiological recovery and sustained attention performance) remains a priority for future research.

4.3.3 Future research

Three research priorities would advance the understanding of urban restorativeness. First and most critical, field-based validation should test whether URPS scores predict actual restoration outcomes. Studies using ecological momentary assessment, pre–post intervention designs, or physiological measures (e.g., heart rate variability or cortisol) would establish whether perceptual ratings translate into measurable recovery. Virtual reality methods could bridge the gap between controlled laboratory conditions and real-world complexity. Second, cross-cultural replication is essential to determine generalizability. Research in diverse cities and cultural contexts would clarify whether the four-factor structure is universal or culture-specific, and whether context-dependent patterns (e.g., quality dominating in parks) reflect universal psychological processes or local norms and expectations. Third, measurement refinement and extension would strengthen the scale's utility. Priorities include establishing test–retest reliability, assessing convergent validity with established restoration scales (e.g., PRS), testing discriminant validity against related constructs (e.g., aesthetic preference and perceived safety), and exploring whether alternative scene classification schemes yield similar patterns. Extensions might include developing short-form versions for rapid assessment or expanded versions capturing setting-specific nuances (e.g., transit-specific wayfinding items).

5 Conclusion

This study developed and validated the Urban Restorative Potential Scale, a theory-driven instrument that assesses restoration potential in everyday urban environments. Psychometric testing with 1,001 participants demonstrated considerable initial reliability, validity, and a four-factor hierarchical structure comprising quality, functionality, captivation, and relaxation. The central finding is that urban restorativeness is context dependent: quality dominates in recreational settings, captivation in housing and commercial areas, and functionality in transit environments. This pattern challenges universal design prescriptions and supports tailored interventions appropriate to each setting type. For practice, the URPS provides a diagnostic tool that identifies specific strengths and weaknesses, making possible evidence-informed design decisions. Theoretically, this work demonstrates that restoration theory extends meaningfully to urban settings when measures capture built environment features alongside natural elements.

By showing that ordinary streets, plazas, and transit hubs can be experienced as more or less restorative depending on their design, this research highlights opportunities to create urban environments that may better support psychological well-being.

Zaldy F. Corpuz, Pangasinan State University Urdaneta Campus, Urdaneta, Philippines
E-mail: zcorpuz@psu.edu.ph

References

Aguinis, H. & Bradley, K. J. (2014) Best-practice recommendations for experimental vignette methodology. *Organizational Research Methods*, 17(4), 351–371. doi:10.1177/1094428114547952

Bornioli, A. & Subiza-Pérez, M. (2023) Restorative urban environments for healthy cities: A theoretical model for the study of restorative experiences in urban built settings. *Landscape Research*, 48(1), 152–163. doi:10.1080/01426397.2022.2124962

Bowler, D., Buyung-Ali, L., Knight, T. & Pullin, A. (2010) A systematic review of evidence for the added benefits to health of exposure to natural environments. *BMC Public Health*, 10(1), 456. doi:10.1186/1471-2458-10-456

Burnham, K. P. & Anderson, D. R. (2002) *Model selection and multimodel inference* (2nd ed.). New York, Springer.

Byrne, B. M. (2016) *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (3rd ed.). New York, Routledge. doi:10.4324/9781315757421

Chen, F. F. (2007) Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14(3), 464–504. doi:10.1080/10705510701301834

Cheung, G. W. & Rensvold, R. B. (2002) Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2), 233–255. doi:10.1207/S15328007SEM0902_5

Cooper Marcus, C. & Sachs, N. A. (2014) *Therapeutic landscapes: An evidence-based approach to designing healing gardens and restorative outdoor spaces*. Hoboken, NJ, Wiley.

Data Privacy Act of 2012. Official Gazette of the Republic of the Philippines, Republic Act no. 10173/2012. Available at: https://www.officialgazette.gov.ph/2012/08/15/republic-act-no-10173/ (accessed 21 Oct. 2025).

Evans, G. W. & Wener, R. E. (2007) Crowding and personal space invasion on the train: Please don't make me sit in the middle. *Journal of Environmental Psychology*, 27(1), 90–94. doi:10.1016/j.jenvp.2006.10.002

Fornell, C. & Larcker, D. F. (1981) Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39–50. doi:10.1177/002224378101800104

Gaskin, J. & Lim, J. (2016) *Model fit measures. Gaskination's StatWiki*. Available at: http://statwiki.gaskination.com (accessed 30 Oct. 2025).

Hair, J. F., Black, W. C., Babin, B. J. & Anderson, R. E. (2010) *Multivariate data analysis* (7th ed.). Upper Saddle River, NJ, Pearson Prentice Hall.

Hartig, T., Korpela, K., Evans, G. W. & Gärling, T. (1997) A measure of restorative quality in environments. *Scandinavian Journal of Psychology*, 37(4), 378–393. doi:10.1080/02815739708730435

Henseler, J., Ringle, C. M. & Sarstedt, M. (2015) A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, 43(1), 115–135. doi:10.1007/s11747-014-0403-8

Hu, L.-t. & Bentler, P. M. (1999) Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3(4), 424–453. doi:10.1037//1082-989X.3.4.424

Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M. & DeShon, R. P. (2012) Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology*, 27(1), 99–114. doi:10.1007/s10869-011-9231-8

Jennings, V. & Bamkole, O. (2019) The relationship between social cohesion and urban green space: An avenue for health promotion. *International Journal of Environmental Research and Public Health*, 16(3), 452. doi:10.3390/ijerph16030452

Joye, Y. & Dewitte, S. (2018) Nature's broken path to restoration: A critical look at Attention Restoration Theory. *Journal of Environmental Psychology,* 59, 1–8. doi:10.1016/j.jenvp.2018.08.006

Joye, Y. & van den Berg, A. E. (2011) Is love for green in our genes? A critical analysis of evolutionary assumptions in restorative environments research. *Urban Forestry & Urban Greening*, 10(4), 261–268. doi:10.1016/j.ufug.2011.07.004

Kaplan, R. & Kaplan, S. (1989) *The experience of nature: A psychological perspective*. New York, Cambridge University Press.

Kaplan, S. (1995) The restorative benefits of nature: Toward an integrative framework. *Journal of Environmental Psychology*, 15(3), 169–182. doi:10.1016/0272-4944(95)90001-2

Karmanov, D. & Hamel, R. (2008) Assessing the restorative potential of contemporary urban environments: Beyond the nature versus urban dichotomy. *Landscape and Urban Planning*, 86(2), 115–125. doi:10.1016/j.landurbplan.2008.01.004

Kline, R. B. (2016) *Principles and practice of structural equation modeling* (4th ed.). New York, Guilford Press.

Korpela, K. M., Ylén, M., Tyrväinen, L. & Silvennoinen, H. (2010) Favorite green, waterside and urban environments, restorative experiences and perceived health in Finland. *Health Promotion International*, 25(2), 200–209. doi:10.1093/heapro/daq007

Li, C.-H. (2016) Confirmatory factor analysis with ordinal data: Comparing estimators. *Behavior Research Methods*, 48(3), 936–949. doi:10.3758/s13428-015-0619-7

Lindal, P. J. & Hartig, T. (2013) Architectural variation, building height, and the restorative quality of urban residential streetscapes. *Journal of Environmental Psychology*, 33, 26–36. doi:10.1016/j.jenvp.2012.09.003

Lindal, P. J. & Hartig, T. (2015) Effects of urban street vegetation on judgments of restoration likelihood. *Urban Forestry & Urban Greening*, 14(2), 200–209. doi:10.1016/j.ufug.2015.02.001

McDonald, R. I., Beatley, T. & Elmqvist, T. (2018) The green soul of the concrete jungle: The urban century, the urban psychological penalty, and the role of nature. *Sustainable Earth*, 1, 3. doi:10.1186/s42055-018-0002-5

Menardo, E., Brondino, M., Damian, O., Lezcano, M., Marossi, C. & Pasini, M. (2024) Students' perceived restorativeness of university environment: The validation of the Rest@U scale. *Frontiers in Psychology*, 15, 1348483. doi:10.3389/fpsyg.2024.1348483

Menardo, E., Brondino, M., Hall, R. & Pasini, M. (2021) Restorativeness in natural and urban environments: A meta-analysis. *Psychological Reports*, 124(2), 417–437. doi:10.1177/0033294119884063

Milfont, T. L. & Fischer, R. (2010) Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research*, 3(1), 111–121. doi:10.21500/20112084.857

Nordh, H., Hartig, T., Hagerhall, C. M. & Fry, G. (2009) Components of small urban parks that predict the possibility for restoration. *Urban Forestry & Urban Greening*, 8(4), 225–235. doi:10.1016/j.ufug.2009.06.003

Norman, G. (2010) Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education*, 15(5), 625–632. doi:10.1007/s10459-010-9222-y

Pan American Health Organization (2020) *Building health throughout the life course: Concepts, implications, and application in public health*. Washington, DC, Pan American Health Organization. Available at: https://iris.paho.org/handle/10665.2/53409 (accessed 20 Oct. 2025).

Pasini, M., Berto, R., Brondino, M., Hall, R. & Ortner, C. (2014) How to measure the restorative quality of environments: The PRS-11. *Procedia – Social and Behavioral Sciences*, 159, 293–297. doi:10.1016/j.sbspro.2014.12.402

Peschardt, K. K. & Stigsdotter, U. K. (2013) Associations between park characteristics and perceived restorativeness of small public urban green spaces. *Landscape and Urban Planning*, 112, 26–39. doi:10.1016/j.landurbplan.2012.12.013

Philippine Health Research Ethics Board (2022) *National ethics guidelines for research involving human participants.* Taguig, Philippine Council for Health Research and Development.

Poulton, E. C. (1982) Biases in quantitative judgements. *Applied Ergonomics,* 13(1), 31–42. doi:10.1016/0003-6870(82)90129-6

Purcell, T., Peron, E. & Berto, R. (2001) Why do preferences differ between scene types? *Environment and Behavior*, 33(1), 93–106. doi:10.1177/00139160121972882

Putnick, D. L. & Bornstein, M. H. (2016) Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, 41, 71–90. doi:10.1016/j.dr.2016.06.004

Rhemtulla, M., Brosseau-Liard, P. É. & Savalei, V. (2012) When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods,* 17(3), 354–373. doi:10.1037/a0029315

Ríos-Rodríguez, M., Rosales, C., Lorenzo, M., Muinos, G. & Hernández, B. (2021) Influence of perceived environmental quality on the perceived restorativeness of public spaces. *Frontiers in Psychology*, 12, 644763. doi:10.3389/fpsyg.2021.644763

San Juan, C., Subiza-Pérez, M. & Vozmediano, L. (2017) Restoration and the city: The role of public urban squares. *Frontiers in Psychology*, 8, 2093. doi:10.3389/fpsyg.2017.02093

Stamps, A. E. (2010) Use of photographs to simulate environments: A meta-analysis. *Perceptual and Motor Skills*, 111(2), 1–30. doi:10.2466/PMS.71.7.907-913

Subiza-Pérez, M., Vozmediano, L. & San Juan, C. (2020) Welcome to your plaza: Assessing the restorative potential of urban squares through survey and objective evaluation methods. *Cities*, 100, 102461. doi:10.1016/j.cities.2019.102461

Ulrich, R. S. (1984) View through a window may influence recovery from surgery. *Science*, 224(4647), 420–421. doi:10.1126/science.6143402

Ulrich, R. S., Simons, R. F., Losito, B. D., Fiorito, E., Miles, M. A. & Zelson, M. (1991) Stress recovery during exposure to natural and urban environments. *Journal of Environmental Psychology*, 11(3), 201–230. doi:10.1016/S0272-4944(05)80184-7

Wilkie, S. & Clements, H. (2018) Further exploration of environment preference and environment type congruence on restoration and perceived restoration potential. *Landscape and Urban Planning*, 170, 314–319. doi:10.1016/j.landurbplan.2017.04.013

Yuan, K.-H. & Bentler, P. M. (2004) On chi-square difference and z tests in mean and covariance structure analysis when the base model is misspecified. *Educational and Psychological Measurement*, 64(5), 737–757. doi:10.1177/0013164404264853

Zhao, J., Wu, J. & Wang, H. (2020) Characteristics of urban streets in relation to perceived restorativeness. *Journal of Exposure Science & Environmental Epidemiology*, 30, 309–319. doi:10.1038/s41370-019-0188-4