Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Design Suitability as the Lesson: Comparison of Research Designs and Mixed-Methods Appraisal in Urban Forestry Teaching Beyond the Classroom

Archana Chaturvedi
Indira Gandhi National Tribal University
Vivek Vaishnav
Vivekvaishnaw@live.in

Manipur University

Research Article

Keywords: research design education, experiential learning, confounding, mixed methods, urban forestry, environmental education

Posted Date: November 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11262277/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Research design is a foundational competence in environmental science education, yet its implications are rarely experienced firsthand before graduate research begins. This paper reports a three-part field exercise conducted with Forestry postgraduate students at Manipur University, India, in which the methodological lesson emerged from students' own data rather than from the classroom's lecture-based instruction. Students first collected beneath-canopy air-quality data under a Completely Randomized Design (CRD) for five canopy categories, including a control (n = 25) and, independently, under a spatially structured campus-quadrant design (n = 60). Because tree-species composition differed across quadrants, the second design did not meet the completeness condition of a true Randomized Block Design (RBD); we treat this shortfall as the central teaching artifact rather than a defect to be hidden. Under the CRD, five of eight air-quality parameters showed large canopy-related effects (partial η^2 = 0.48–0.75); once spatial block was accounted for, most of these effects weakened substantially, with block itself becoming the significant source of variance for particulate matter. A second exercise asked 25 students from 25 departments to complete a short perception survey; 88% reported that campus greenery positively influenced their academic life, a dimension that the biophysical measurements could not capture. A third exercise required students to compare all findings and to judge, rather than assume, what each design could and could not support. We describe the exercises, the assessment rubric used, and the broader case for teaching design suitability as an explicit, examinable skill in biology and environmental-science curricula.

Introduction

Environmental science education is expected to produce graduates who can generate credible evidence, not merely recite the vocabulary of experimental design. Among the competencies, such training must build the ability to judge whether a given design can actually support the conclusion being drawn from it. This judgment precedes data collection and shapes what can legitimately be reported afterward. In practice, research design is often taught as a set of labels to be memorized: students can define a Completely Randomized Design (CRD) and a Randomized Block Design (RBD) and recite their assumptions, yet rarely meet a situation in their own coursework where a design's structural limits, rather than its textbook description, determine what can be concluded through the observations beyond the classrooms.

This paper describes a compact intervention beyond the classroom, centered on three linked practical exercises conducted by Forestry postgraduate students during a research methodology module at Manipur University, Imphal. The setting is urban forestry, a topic of obvious relevance in rapidly urbanizing towns and cities, where claims about which tree species best mitigate air pollution are only as trustworthy as the design used to isolate species identity from the site conditions in which species happen to be planted (Kaur & Nagpal, 2017; Prajapati & Tripathi, 2008). Rather than engineering a textbook-clean comparison, we let students discover, through their own data, that a design nominally described as "blocked" does not automatically deliver the inferential protection of a true block design,

and that recognizing this gap is itself the more advanced skill. The exercise sits within a wider conversation about how practical, technology- and data-rich activities beyond the classroom can be used to build methodological competence rather than deliver content. This work extends to the logically prior question of experimental design: before students can trust what a tool or dataset tells them about the natural world, they need practice recognizing when the structure of their own study can and cannot support that inference.

The intervention had three learning objectives. First, students would learn to evaluate whether a design structure suits the question being asked of it, rather than treating a design label as a guarantee of validity. Second, students would recognize that quantitative environmental monitoring and qualitative social appraisal answer different, complementary questions. Third, students would build practical competence across the full research cycle: collection, analysis, critical evaluation of design validity, and integration across methods. The broader aim is a transferable teaching model, grounded in experiential learning theory and place-based science education, in which the most consequential lesson that a design must be judged fit for purpose before its results are trusted comes from the students' own data.

Theoretical Framework

Experiential learning and the teaching of research methods

The design draws on experiential learning theory, in which learning proceeds through a cycle of concrete experience, reflective observation, abstract conceptualization, and active experimentation (Kolb & Kolb, 2005). Applied to research-methods teaching, this implies that students should not simply be told about the limitations of a design; they should meet those limitations in data they collected themselves and be asked to judge, rather than assume, whether the design supports the claim being made from it. Reviews of statistical practice in educational research document widespread shortfalls, infrequent checking of design assumptions, sparse reporting of effect sizes, and limited attention to power (Keselman et al., 1998), which we read as symptomatic of design being taught as a procedure to execute rather than an assumption to interrogate. Guidance on planning comparative experiments in educational settings likewise stresses that identifying a design's salient pitfalls, not just following its steps, is the analytical skill worth teaching (McGowan, 2011). Place-based science education (Semken & Freeman, 2008) and the wider inquiry-learning literature (Pedaste et al., 2015) support locating this lesson on the students' own campus, where its consequences are immediately visible.

Design suitability and incomplete blocking

A canonical RBD requires every treatment level to appear, ideally with equal replication, within every block. When this condition fails, for instance, when a putative treatment (here, tree species) is distributed unevenly across blocks, treatment and block effects become partially confounded, and the resulting F-tests no longer isolate a "pure" treatment effect in the way a textbook RBD would isolate (Hurlbert,

1984). Hurlbert's classic analysis of pseudoreplication in ecological field experiments is instructive precisely because it documents how often field researchers apply inferential statistics to designs whose replication structure does not, in fact, support the error term used and argues that recognizing this mismatch is a core design-literacy skill, not a technical footnote. Where species and location are confounded, as in our campus dataset, neither a naïve CRD nor a naïve "blocked" analysis that treats species as a clean, replicated factor can support a species-specific claim. Our pedagogical response is not to force the data into a corrected RBD, but to have students recognize the confound, de-identify species as the object of inference, and instead judge what the realized design can and cannot support: a comparison between canopied and open conditions across blocks, and a judgment about whether location or an unreplicated species factor better explains the variance. This reframing does not dismiss the substantive literature on species-specific air-pollution tolerance (Kaur & Nagpal, 2017; Prajapati & Tripathi, 2008); it simply insists that such claims are only as credible as the design that isolates species from confounding site variables.

Complementary quantitative and qualitative inquiry

The third exercise, a short participatory questionnaire, introduced students to the complementarity of quantitative and qualitative modes of inquiry. Environmental problems are socio-ecological, and integrating biophysical and perception-based evidence is increasingly regarded as necessary for validity in environmental research (Fetters et al., 2013; Clark, 2019). Having students design, administer, and interpret their own questionnaire rather than analyze a pre-collected dataset, extended to the qualitative component, provides the same first-hand ownership. Exercise 1 gave the quantitative component, consistent with evaluation frameworks for citizen-science learning that stress depth and structure of participation over data collection alone (Jordan et al., 2012; Phillips et al., 2018), and with classroom studies reporting gains in interest and self-efficacy from structured participatory assignments (Cotey et al., 2025; Peter et al., 2019; Pizzolato & Tsuji, 2022).

From method anxiety to method competence

A further strand of literature concerns the well-documented anxiety many students feel when first encountering research design and statistics, associated with self-perception as much as ability, with case-based, real-data instruction repeatedly identified as a mitigating strategy (Alladin et al., 2026; Papanastasiou & Zembylas, 2008; Chew & Dillon, 2014). Course-based undergraduate research experiences (CUREs), in which whole classes work on a genuine, unresolved research question using their own data, have been shown to improve research self-efficacy and retention outcomes at scale (Rodenbusch, et al., 2016), including in field-based ecological settings closely analogous to the present study (Stanfield et al., 2022). The intervention reported here can be read as a compact, single-module CURE: it substitutes a genuine, unresolved design problem, whether canopy category or spatial block better explains the observed variance, for a pre-scripted laboratory exercise with a known answer.

Context and Setting

The exercise was conducted on the campus of Manipur University (24.7509° N, 93.9302° E), Imphal, Northeast India (approximately 287 hectares), which supports a diverse assemblage of avenue and roadside tree plantations under a subtropical climate with distinct wet and dry seasons. Avenue plantings include native species such as *Mangifera indica*, *Lagerstroemia speciosa*, and *Tectona grandis*, as well as introduced ornamentals such as *Callistemon citrinus* and *Kigelia africana*. The campus was chosen for three reasons: sufficient size and tree diversity for a comparative study; a natural division into four quadrants (northwest, northeast, southeast, southwest) that captured real spatial variation in tree distribution and proximity to roads, offering a candidate blocking factor whose realized species distribution, as described below, did not in fact satisfy the completeness a formal RBD requires; and an accessible, disciplinarily diverse student population for the qualitative component. Participating students were enrolled in the Forestry postgraduate program. They had some prior exposure to basic statistics but limited practical experience with how design choices shape what can be inferred, including the structural limitations involved. Fieldwork ran for approximately four weeks as part of a four-credit research methodology module.

The questionnaire component involved voluntary, self-reported data (including age and gender) from fellow students, who were informed of the purpose of data collection, given consent, and assured that responses would be reported only in an aggregate, anonymized form; no individually identifying information appears in this paper. The exercise was conducted under departmental oversight as part of the credit-bearing module.

The Three Exercises

Exercise 1: comparative design for air-quality assessment

Students were asked to compare an unblocked (CRD) and a spatially structured design for evaluating tree-canopy categories, with the explicit hypothesis that the two designs would yield different conclusions about canopy-related air-quality effects because of their different replication structures. For the CRD, five canopy categories were sampled, including four tree species: *Kigelia africana, Lagerstroemia speciosa, Callistemon citrinus, Enterolobium cyclocarpum*, and an open-space control, with five replicate trees each (n = 25). For each tree, students recorded height, canopy diameters (two perpendicular directions), and beneath-canopy air quality: formaldehyde (HCHO), total volatile organic compounds (TVOC), particulate matter in three size fractions (PM2.5, PM1.0, PM10.0), temperature, relative humidity, and a computed Air Pollution Index (API), using a Portable Air Quality Pollution Monitor Meter.

For the spatially structured component, the campus was divided into four quadrants; within each, three replicate trees per locally available species were sampled, with species composition varying naturally across quadrants (60 observations across nine species and a control). Because most species occurred

in only one or two of the four quadrants, this component does not meet the completeness condition of a true RBD with respect to species identity. We therefore analyze canopy category as a de-identified categorical variable (spatial canopy position rather than species label) and treat the mismatch between the intended blocking structure and the realized species distribution as an explicit limitation and teaching point, not as a validated species-blocked comparison.

Exercise 2: participatory appraisal of campus-greenery perceptions

A structured questionnaire was administered to 25 students from 25 academic departments, eliciting perceptions of campus biodiversity, the influence of greenery on academic life, the most valued environmental features, comparative assessments with other universities, ranked environmental priorities, and willingness to join student-led environmental initiatives.

Exercise 3: analysis, design evaluation, and integration

Students performed descriptive and inferential statistics on the Exercise 1 and 2 data. They were explicitly asked to evaluate, rather than assume, the suitability of each design for the claim it supports. For the CRD data, they ran a one-way ANOVA with canopy category as the factor; for the spatially structured data, they ran a two-way ANOVA with canopy category and block, computing effect sizes throughout and identifying where incomplete blocking limited what could be inferred about species identity specifically, as distinct from canopy presence or spatial location generally. For the qualitative data, they computed frequency distributions and cross-tabulations. The central task was not simply to compare CRD and spatially structured results, but to judge which, if either, design actually supported the causal claim in question.

Methods of the Pedagogical Study

Data collection

Air-quality data were collected with a portable multi-parameter monitor (PM2.5, PM1.0, PM10.0, HCHO, TVOC) at a standardized height of 1.5 m, midway between the trunk and the canopy edge, after a two-minute stabilization period, with temperature and humidity recorded simultaneously. Because measurements were not stratified by time of day or replicated across seasons, diurnal and seasonal variability could not be disentangled from canopy-category or block effects — a constraint revisited in the Limitations. Control measurements were taken at open sites within each block. Questionnaire data were collected from 25 students spanning natural sciences, social sciences, humanities, and professional fields.

Statistical analysis

One-way ANOVA was used for the CRD data (canopy category, five levels); two-way ANOVA (Type II sums of squares) for the spatially structured data (canopy category and block, four levels), on the assumption

that the category-by-block interaction was not the primary object of inference, given limited replication per cell. Analyses were performed using Python (SciPy, Statsmodels). Significance was assessed at $\alpha = 0.05$, with a Bonferroni-adjusted threshold of $\alpha = 0.00625$ (0.05/8) reported alongside it because eight parameters were tested, and partial eta-squared (η^2_p) was reported for every F-test following standard effect-size practice for ANOVA designs (Lakens, 2013). Reporting effect sizes lets a non-significant result, expected given the modest sample size of a teaching exercise, still be recognized as practically large, a distinction we wanted students to encounter directly (Keselman et al., 1998; Lakens, 2013).

Assessing student learning

Learning was assessed from students' submitted deliverables, descriptive tables, ANOVA summaries with effect sizes, graphics, and written discussion, against the four-criterion rubric in Table 1, which treats design-suitability judgment as a distinct, separately graded criterion so that a student who executed the analysis correctly but missed the incomplete blocking would not be scored as demonstrating full design literacy. This qualitative assessment of student work, while not a formal pre–post design with a comparison group, gives evidence of the depth of understanding reached; we note the absence of a controlled evaluation as a limitation and recommend it for future iterations.

Table 1
Assessment rubric used to evaluate student deliverables from Exercises 1–3

Criterion	What deliverables were evaluated against
Statistical execution	Correct computation and reporting of descriptive statistics and ANOVA results, including degrees of freedom and, where applicable, effect sizes.
Mechanism identification	Ability to articulate, in writing, why the CRD and spatially structured analyses produced different conclusions (i.e., recognition of confounding between species and location).
Design-suitability judgment	Ability to assess whether a given design (CRD, RBD, or the incomplete block structure actually realized) was suitable for the question asked, rather than treating any design as automatically valid.
Cross-paradigm integration	Ability to compare and integrate quantitative and qualitative findings in a three-way comparison.

Findings

CRD results: apparent canopy-category effects

Under the CRD, one-way ANOVA showed statistically significant differences among canopy categories for several parameters, the largest for PM2.5 ($F(4, 20) = 10.58$, $p < .001$, $\eta^2_p = 0.68$), relative humidity ($F(4, 20) = 15.01$, $p < .001$, $\eta^2_p = 0.75$), and the Air Pollution Index ($F(4, 20) = 13.09$, $p < .001$, $\eta^2_p = 0.72$). Significant effects were also found for PM1.0, PM10.0, and temperature; only HCHO and TVOC were non-significant. Descriptive statistics are in Table 2, and the full comparison, including effect sizes, is in Table 3.

Table 2
Descriptive statistics (mean ± SD) for air-quality parameters under the CRD

Category	PM2.5	PM10.0	TVOC	Temp (°C)	Humidity (%)	API
Control (open)	4.80 ± 1.79	5.00 ± 2.24	0.053 ± 0.031	31.6 ± 1.34	54.0 ± 5.70	5.0 ± 1.58
Kigelia africana	3.80 ± 1.10	3.80 ± 1.10	0.081 ± 0.029	28.2 ± 1.30	66.6 ± 7.09	1.2 ± 0.84
Lagerstroemia speciosa	3.40 ± 0.55	3.80 ± 0.84	0.074 ± 0.010	28.6 ± 1.52	63.6 ± 2.19	1.6 ± 0.55
Callistemon citrinus	8.40 ± 2.07	10.60 ± 5.03	0.086 ± 0.032	27.6 ± 0.89	72.2 ± 2.17	2.6 ± 0.89
Enterolobium cyclocarpum	5.20 ± 0.45	5.20 ± 0.45	0.096 ± 0.006	27.4 ± 1.14	72.6 ± 2.07	2.0 ± 0.00

Note: PM values in μg/m³; TVOC in mg/m³; Temp in °C; Humidity in %; API = Air Pollution Index. n = 5 per canopy category.

What the spatially structured analysis can and cannot support

With canopy category de-identified from species and treated as a categorical spatial position, the canopy-category effect for particulate matter attenuated substantially relative to the CRD: PM2.5 ($F(9, 41) = 1.93$, $p = .071$, $\eta^2p = 0.30$), PM1.0 ($F = 1.93$, $p = .071$, $\eta^2p = 0.30$), PM10.0 ($F = 1.77$, $p = .099$, $\eta^2p = 0.28$). Block was statistically significant for all three particulate fractions: PM2.5 ($F(3, 41) = 3.45$, $p = .024$), PM1.0 ($F = 3.47$, $p = .023$), PM10.0 ($F = 3.13$, $p = .035$).

It is worth stating plainly what this does and does not license. It does not show that species identity does not affect particulate matter; that would require species to be replicated across all blocks, a condition this configuration does not provide. What it does show is a design-level observation: when canopy category and spatial block are confounded, a substantial share of the variance that a naïve CRD analysis attributes to canopy category is at least as plausibly attributable to location. The lesson for students is about the design, not the trees: an incomplete block structure of this kind cannot adjudicate between a species explanation and a location explanation, and any published species-specific recommendation drawn from a similarly confounded design should be treated with caution.

Relative humidity was the one parameter whose canopy-category effect remained significant under both structures (CRD: $F = 15.01$, $p < .001$, $\eta^2p = 0.75$; spatially structured: $F = 7.01$, $p < .001$, $\eta^2p = 0.61$), with a non-significant block effect ($F = 1.56$, $p = .212$, $\eta^2p = 0.10$), the one parameter for which a canopy-related explanation remains plausible without being ruled out by the location confound. However, attributing it specifically to species rather than canopy presence in general would still require a properly replicated design. Temperature, by contrast, was significant under the CRD ($F = 9.15$, $p < .001$, $\eta^2p = 0.65$) but non-

significant for both canopy category (F = 1.22, p = 0.306, η^2p = 0.21) and block (F = 0.42, p = 0.743, η^2p = 0.03) under the spatially structured analysis, a useful demonstration that a significant CRD result is not self-validating (Table 3).

Effect sizes and the limits of significance testing

Several parameters with non-significant canopy-category effects under the spatially structured analysis nonetheless carried large partial eta-squared values, PM2.5 and API each retained $\eta^2p \approx 0.30$ despite $p > 0.05$. This is consistent with the small cell sizes inherent to a teaching-scale exercise and gives students a concrete encounter with the distinction between statistical non-significance and the absence of a practically meaningful effect (Keselman et al., 1998; Lakens, 2013). Students were asked to reconcile each such non-significant p-value with its moderate-to-large effect size and to argue, in writing, why low statistical power is the more defensible interpretation rather than the absence of an underlying effect.

Table 3
Comparison of ANOVA results, with effect sizes, under the CRD and the spatially structured (incomplete block) design

Parameter	CRD F (p)	CRD η^2p	Canopy F (p)	Canopy η^2p	Block F (p)	Block η^2p
PM2.5	10.58 (< 0.001)***	0.68	1.93 (0.071) ns	0.30	3.45 (0.024)*	0.20
PM1.0	4.68 (0.008)**†	0.48	1.93 (0.071) ns	0.30	3.47 (0.023)*	0.20
PM10.0	6.17 (0.002)**	0.55	1.77 (0.099) ns	0.28	3.13 (0.035)*	0.19
HCHO	2.11 (0.118) ns	0.30	1.36 (0.236) ns	0.23	2.72 (0.055) ns	0.17
TVOC	2.21 (0.105) ns	0.31	2.31 (0.030)*	0.34	3.03 (0.038)*	0.18
Temperature	9.15 (< 0.001)***	0.65	1.22 (0.306) ns	0.21	0.42 (0.743) ns	0.03
Humidity	15.01 (< 0.001)***	0.75	7.01 (< 0.001)***	0.61	1.56 (0.212) ns	0.10
API	13.09 (< 0.001)***	0.72	1.97 (0.065) ns	0.30	3.16 (0.033)*	0.19

Note: CRD df = 4, 20; spatially structured canopy-category df = 9, 41; block df = 3, 41. η^2p computed from reported F and df (Lakens, 2013). *p < 0.05; **p < 0.01; ***p < 0.001; ns = not significant. †PM1.0 (p = 0.008) does not survive a Bonferroni-corrected threshold of $\alpha = 0.00625$; PM2.5, PM10.0, temperature, humidity, and API remain significant under this stricter threshold. Canopy category is de-identified from species.

Qualitative findings: student perceptions of campus greenery

The survey reached 25 students from 25 departments, aged 20–32 (mean = 23.4; 19 male, 4 female, 2 other genders; 23 of 25 were postgraduate). The disciplinary spread was deliberate; the gender composition was not balanced by design, and the percentages below should be read as descriptive of this convenience sample rather than as generalizable estimates. Fourteen respondents (56%) rated campus biodiversity as moderate, 10 (40%) as high, and 1 (4%) as low. Twenty-two (88%) agreed or strongly agreed that greenery positively influenced academic life; 3 (12%) were neutral; none disagreed. Trees and plants were the most valued campus feature (52%), followed by clean air and shade (28%), gardens and lawns (8%), and wildlife (8%). 12 (48%) rated the campus ecological condition as somewhat better than at other universities, 8 (32%) much better, 4 (16%) about the same, and 1 (4%) did not know. Shade trees were the top-ranked priority (36%), followed by water management and cleanliness (32%), noise and pollution control (20%), and flower gardens (8%). 18 (72%) were willing to join student-led environmental initiatives, 6 (24%) said maybe, and 1 (4%) said no.

Beyond its substantive content, tabulating these perceptions confronted student researchers with the range of values different members of the campus community attach to urban greenery, complicating any assumption of uniform appreciation. The pattern is broadly consistent with the wider literature linking student use of green space to perceived quality of life (McFarland et al., 2008) and campus environment to student wellbeing more generally (Ding et al., 2023), with survey evidence elsewhere finding students perceive urban green space as beneficial across physical, mental, and social dimensions (Saha & Haq, 2024), and with links between green infrastructure, community cohesion (Weinstein et al., 2015), social interaction (Peters, Elands, & Buijs, 2010), and restorative wellbeing (Akpinar, 2021; Fisher et al., 2021). This supports treating the qualitative survey not as an add-on but as tapping a dimension of urban-forestry value that biophysical measurement alone cannot reach.

The pivotal comparison: design, not species, as the object of learning

The central pedagogical moment came when students compared the CRD and spatially structured analyses side by side and wrote a rationale for every parameter. Five of the eight air-quality parameters significant under the CRD became non-significant for the canopy category under the spatially structured analysis. In contrast, block became significant for the particulate matter parameters. Students were asked to articulate, in design terms, the mechanism that confounded the canopy category. Spatial block meant the CRD's apparent effect could not be cleanly separated from a location effect. It was explicitly guided away from concluding that the spatially structured analysis had "proven" species did not matter, towards the more defensible conclusion that the available data could not adjudicate between a species explanation and a location explanation. Recognizing that limit, rather than reporting either design's result at face value, was itself the competence being assessed.

The three-way comparison

Students compared findings across all three exercises and summarized what each contributed (Table 4). The CRD suggested canopy category mattered for beneath-canopy air quality; the spatially structured

analysis showed this could not be cleanly attributed to species once an unreplicated location confound was acknowledged; and the survey showed that 88% of students perceived campus greenery as beneficial to academic life, with shade trees the most valued priority, a dimension neither quantitative design could reach. Only by comparing all three could students arrive at a picture of what the campus urban forest means, what its air-quality effects plausibly are given the design's limits, and what it could become.

Table 4
Three-way comparison of exercises, with the spatially structured design reframed around design suitability

Dimension	Exercise 1: CRD	Exercise 1: Spatially structured RBD	Exercise 2: Qualitative survey
Research question	Do tree canopies differ in beneath-canopy air quality?	Does a spatial block better explain the apparent canopy effect than an unreplicated species factor?	How do people perceive and value campus greenery?
Paradigm	Quantitative, experimental, unblocked	Quantitative, spatially structured, but incompletely blocked for species	Qualitative, participatory, interpretive
Key finding	Apparent canopy-category effects (e.g., PM2.5: $p < .001$)	Canopy-category effect weakens ($p = .071$); block emerges as significant	88% perceive greenery as beneficial; shade trees are most valued
Implication	Risk of recommending species without ruling out site confounding	Prioritize rigorous site-level design before attributing outcomes to species	Engage community; trees valued for shade, air quality, and well-being
What it cannot tell	Whether people value the trees, whether spatial confounding is present	Reliable species-specific effects, because species were not replicated across blocks	Whether trees actually alter measured air quality
Note: No single approach provides a complete or unconfounded account of the campus urban forest.			

Discussion

Design suitability as an epistemic choice

The findings underscore a point well established in the statistical literature but insufficiently internalized in biology and environmental-science teaching (Reiss, 2016): the structure of a research design is not a procedural detail but an epistemic choice that determines what can be known, and its suitability for a given claim must be actively assessed rather than assumed from its textbook label (Hurlbert, 1984). The CRD, by treating all experimental units as exchangeable, assumes no systematic source of variability other than the treatment exists; when this is violated, as it almost always is in field-based environmental research, the design can produce substantively misleading and statistically significant results. A spatially

structured design only delivers on its promise of separating treatment from location when the treatment is in fact replicated across blocks, a condition our dataset did not meet for species identity. Recognizing that shortfall, rather than reporting the analysis as if the condition had been met, is the more advanced competence. For students preparing their own thesis research, the practical warning is that the significance of a result may depend less on the strength of an underlying phenomenon than on whether their design's replication structure actually supports the comparison being drawn. That part of competent practice is being able to say, in writing, when it does not.

Implications for urban-forestry research and teaching

A large literature quantifies species-specific differences in air-pollution tolerance, typically derived from leaf-level biochemical traits (Keller & Schwager, 1977; Nadgórska-Socha, et al., 2017; Ogunkunle, et al., 2015; Prajapati & Tripathi, 2008), and field applications consistently report meaningful interspecific variation in tolerance and green-belt suitability (Kaur & Nagpal, 2017; Nkansah, et al., 2025; Pandey, et al., 2015; Pathak, et al., 2011), while city-scale assessments confirm real aggregate air-quality benefits from urban tree cover (Nowak, et al., 2014). Many such studies sample species across sites that differ in traffic exposure or building density without a design that isolates species from location, precisely the confound this exercise reproduces. This is not to say species do not matter: the humidity result, robust under both structures, is consistent with process-level evidence that transpiring canopies reliably raise near-surface humidity and moderate temperature largely independent of species identity (Feng, et al., 2023; Ibsen et al., 2024; Meili et al., 2021). The broader point for teaching is that the reliability of species-specific claims depends on whether a design achieves complete, balanced replication of species across spatial strata, and students who go on to conduct such research need practice recognizing this dependency in their own and others' work.

Complementarity of quantitative and qualitative evidence

Comparing all three exercises reinforced a further principle: different research approaches answer different questions about the same phenomenon, and integrating them meaningfully is itself a skill requiring deliberate design (Clark, 2019). A campus management plan informed only by the CRD would risk selecting species on a possibly confounded ranking; one informed only by the spatially structured analysis, taken at face value, would risk concluding wrongly that species do not matter at all; one informed only by the survey would respond to perceived value without checking whether measurable biophysical benefits are present. Only by integrating all three can urban-forestry decisions be both ecologically sound and socially legitimate.

Transferability

The exercise transfers to other classroom contexts with minimal adaptation. The transferable element is not the content but the structure: students collect real data under two contrasting quantitative structures and one qualitative approach, analyze all three, and are required to assess whether each design's replication structure supports the inference being drawn, before integrating quantitative and qualitative

evidence, rather than assuming. This structure applies equally to water-quality assessment, soil analysis, biodiversity monitoring, or microclimate studies, and connects to the growing interest in the activities beyond the classrooms, built around student-generated, tool-mediated data, for example, exercises that ask students to work with digital identification technologies rather than pre-verified species lists. Because incomplete blocking is common rather than exceptional in field-based student projects, we suggest that deliberately retaining an imperfectly blocked design as a teaching artifact and requiring students to diagnose, rather than paper over, its limitations, is a more realistic and more transferable version of this exercise than one engineered to produce a textbook-clean design.

Learning Outcomes and Implications for Practice

Evidenced in students' submitted work and evaluated against the rubric (Table 1), students demonstrated (a) an understanding that a design's structure, not merely its label, determines what conclusions it can support, shown in written discussions correctly identifying the canopy-by-block confound and its consequences; (b) recognition that quantitative monitoring and qualitative appraisal are complementary modes of inquiry, shown in their three-way comparisons; and (c) practical competence across the research cycle, including the previously under-emphasized skill of evaluating a design's suitability before trusting its output.

For instructors adapting a similar exercise: introduce blocking, confounding, and effect size before fieldwork begins, but do not tell students in advance whether their realized design will achieve complete blocking. Having students discover the incompleteness themselves and respond to it analytically appears to be where the deepest learning occurs. Grade design-suitability judgment explicitly and separately (as in Table 1) rather than treating a correctly executed ANOVA as sufficient evidence of design literacy. Where feasible, add a brief pre- and post-exercise self-assessment of design-literacy concepts to strengthen the evidence base for claimed learning gains, a step not implemented here and flagged as a priority for future replications.

Limitations

Several limitations of this investigation should be acknowledged, foremost the incomplete blocking structure discussed throughout. Because species composition varied across the four campus quadrants, most species occurred in only one or two blocks, so species and block are partially confounded, and the reduced residual degrees of freedom (df = 41, across nine canopy categories and four blocks with 60 observations) further limit the precision of any single comparison. We have addressed this by de-identifying species from the canopy-category factor and restricting substantive claims to the design level rather than the species level; we regard this as the single most important limitation of the underlying dataset and present it as a limitation to be disclosed, not an error to be silently corrected.

Second, learning outcomes were assessed from student deliverables against a rubric rather than a formalized pre–post design with a comparison group; a more rigorous evaluation, incorporating pre- and post-exercise measures of design-literacy concepts, would strengthen future claims about effectiveness. Third, sample sizes were modest by the standards of a substantive study (CRD n = 25; spatially

structured n = 60, with category composition varying across blocks), adequate to demonstrate the design comparison but limiting statistical power and generalizability, a limitation partially addressed by reporting effect sizes alongside *p*-values. Fourth, the survey reached 25 respondents from 25 departments, with diverse disciplinary representation but not a representative or gender-balanced sample; its percentages should be read descriptively. Fifth, the exercise was conducted at a single institution with a specific set of species and conditions, and air-quality measurements were taken at a single time point rather than being replicated seasonally or diurnally, limiting the ability to separate temporal variability from canopy-category or block effects. Despite these limits, the exercise's primary purpose was pedagogical rather than substantive, and its central learning outcome, the capacity to recognize and articulate a design's limitations rather than accept its output uncritically, was achieved regardless of the generalizability of the underlying environmental data.

Conclusion

This paper has described three linked exercises out of the classroom that use a Completely Randomized Design and a spatially structured, incompletely blocked design, together with a participatory appraisal survey, to teach postgraduate Forestry students that a research design's suitability for a given question is a matter of active judgment, not a property guaranteed by its textbook label. The same air-quality dataset, analyzed under two structures, yielded different apparent conclusions about canopy-category effects on particulate matter; a design-suitability-oriented reading showed that neither structure alone could adjudicate between a species explanation and a spatial explanation, given the incomplete replication of species across blocks; and the survey showed that the social value of campus greenery, affirmed by 88% of respondents, was a dimension of urban forestry entirely invisible to the biophysical measurements. By comparing all three results and by being explicit about what each design could and could not support, students discovered that experimental design and qualitative appraisal answer fundamentally different questions about the same urban forest, and that judging a design's suitability, including recognizing its incompleteness, is itself a core research competency worth teaching explicitly in environmental science curricula.

Declarations

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Competing interest

The authors declare that they have no competing interests.

Data availability

Data will be provided on request to the corresponding author.

Author contributions

AC prepared the initial draft and performed the initial data analysis. VV conceptualized and supervised the study and prepared the final draft of the manuscript. All authors read and approved the final manuscript.

Acknowledgment

The authors gratefully acknowledge the institutional support of the Indira Gandhi National Tribal University and Manipur University.

Ethical Approval Declaration

The Research Ethics Committee of the Department of Forestry, Manipur University, approved the methodology and issued the 'Category A' exception certificate. Prior to conducting the field survey, we obtained consent from all the informants. Informants were thoroughly briefed on the study's objectives before the interviews, and verbal consent was obtained and subsequently documented, in accordance with the ISE Code of Ethics and guidelines.

Clinical trial number: not applicable

References

1. Akpinar A (2021) How perceived sensory dimensions of urban green spaces are associated with teenagers' perceived restoration, stress, and mental health. Landsc Urban Plann 214 Article 104185. https://doi.org/10.1016/j.landurbplan.2021.104185
2. Alladin FM, Esnard TR, Samlal KC (2026) The association between utility of learning resources, class attendance, statistical self-efficacy, mathematics competence and statistics anxiety among undergraduate students. J Stat Data Sci Educ. https://doi.org/10.1080/26939169.2026.2616504
3. Chew PKH, Dillon DB (2014) Statistics anxiety update: Refining the construct and recommendations for a new research agenda. Perspect Psychol Sci 9(2):196–208. https://doi.org/10.1177/1745691613518077
4. Clark VLP (2019) Meaningful integration within mixed methods studies: Identifying why, what, when, and how. Contemp Educ Psychol 57:106–111. https://doi.org/10.1016/j.cedpsych.2019.01.007
5. Cotey SR, Dunbar-Wallis AK, Golumbic YN, Mehrotra P, Beamer DA, Vance-Chalcraft HD (2025) Online Citizen Science in Higher Education Courses. Ecosphere 16(7):e70352. https://doi.org/10.1002/ecs2.70352
6. Ding Y, Lee C, Chen X, Song Y, Newman G, Lee R, Lee S, Li D, Sohn W (2023) Exploring the association between campus environment of higher education and student health: A systematic

review of findings and measures, vol 91., Article 128168. Urban Forestry & Urban Greening. https://doi.org/10.1016/j.ufug.2023.128168

7. Feng X, Wen H, He M, Xiao Y (2023) Microclimate effects and influential mechanisms of four urban tree species underneath the canopy in hot and humid areas. Front Environ Sci 11:1108002. https://doi.org/10.3389/fenvs.2023.1108002
8. Fetters MD, Curry LA, Creswell JW (2013) Achieving integration in mixed methods designs—Principles and practices. Health Serv Res 48(6pt2):2134–2156. https://doi.org/10.1111/1475-6773.12117
9. Fisher JC, Irvine KN, Bicknell JE, Hayes WM, Fernandes D, Mistry J, Davies ZG (2021) Perceived biodiversity, sound, naturalness and safety enhance the restorative quality and wellbeing benefits of green and blue space in a neotropical city. Sci Total Environ 755 Article 143095. https://doi.org/10.1016/j.scitotenv.2020.143095
10. Hurlbert SH (1984) Pseudoreplication and the design of ecological field experiments. Ecol Monogr 54(2):187–211. https://doi.org/10.2307/1942661
11. Ibsen PC, Crawford B, Corro LM, Bagstad KJ, McNellis BE, Jenerette GD, Diffendorfer J (2024) Urban tree cover provides consistent mitigation of extreme heat in arid but not humid cities. Sustainable Cities Soc 113:105677. https://doi.org/10.1016/j.scs.2024.105677
12. Jordan RC, Ballard HL, Phillips TB (2012) Key issues and new approaches for evaluating citizen-science learning outcomes. Front Ecol Environ 10(6):307–309. https://doi.org/10.1890/110280
13. Kaur M, Nagpal AK (2017) Evaluation of air pollution tolerance index and anticipated performance index of plants and their application in the development of green space along the urban areas. Environ Sci Pollut Res 24(23):18881–18895. https://doi.org/10.1007/s11356-017-9500-9
14. Keller T, Schwager H (1977) Air pollution and ascorbic acid. Eur J For Pathol 7(6):338–350. https://doi.org/10.1111/j.1439-0329.1977.tb00603.x
15. Keselman HJ, Huberty CJ, Lix LM, Olejnik S, Cribbie RA, Donahue B, Kowalchuk RK, Lowman LL, Petoskey MD, Keselman JC, Levin JR (1998) Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA, and ANCOVA analyses. Rev Educ Res 68(3):350–386. https://doi.org/10.3102/00346543068003350
16. Kolb AY, Kolb DA (2005) Learning styles and learning spaces: Enhancing experiential learning in higher education. Acad Manage Learn Educ 4(2):193–212. https://doi.org/10.5465/amle.2005.17268566
17. Lakens D (2013) Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. Front Psychol 4., Article 863. https://doi.org/10.3389/fpsyg.2013.00863
18. McFarland AL, Waliczek TM, Zajicek JM (2008) The relationship between student use of campus green spaces and perceptions of quality of life. HortTechnology 18(2):232–238. https://doi.org/10.21273/HORTTECH.18.2.232

19. McGowan HM (2011) Planning a comparative experiment in educational settings. J Stat Educ 19(2). https://doi.org/10.1080/10691898.2011.11889612
20. Meili N, Manoli G, Burlando P, Carmeliet J, Chow WTL, Coutts AM, Roth M, Velasco E, Vivoni ER, Fatichi S (2021) Tree effects on urban microclimate: Diurnal, seasonal, and climatic temperature differences explained by separating radiation, evapotranspiration, and roughness effects. Urban Forestry & Urban Greening, 58, Article 126970 https://doi.org/10.1016/j.ufug.2020.126970
21. Nadgórska-Socha A, Kandziora-Ciupa M, Trzęsicki M, Barczyk G (2017) Air pollution tolerance index and heavy metal bioaccumulation in selected plant species from urban biotopes. Chemosphere 183:471–482. https://doi.org/10.1016/j.chemosphere.2017.05.128
22. Nkansah FK, Belford EJD, Hogarh JN (2025) Assessing the air pollution mitigation potential of urban trees in Ghana's Central Region. https://doi.org/10.1155/tswj/9655268. The Scientific World Journal
23. Nowak DJ, Hirabayashi S, Bodine A, Greenfield E (2014) Tree and forest effects on air quality and human health in the United States. Environ Pollut 193:119–129. https://doi.org/10.1016/j.envpol.2014.05.028
24. Ogunkunle CO, Suleiman LB, Oyedeji S, Awotoye OO, Fatoba PO (2015) Assessing the air pollution tolerance index and anticipated performance index of some tree species for biomonitoring environmental health. Agroforest Syst 89(3):447–454. https://doi.org/10.1007/s10457-014-9781-7
25. Pandey AK, Pandey M, Mishra A, Tiwary SM, Tripathi BD (2015) Air pollution tolerance index and anticipated performance index of some plant species for development of urban forest. Urban Forestry Urban Green 14(4):866–871. https://doi.org/10.1016/j.ufug.2015.08.001
26. Papanastasiou EC, Zembylas M (2008) Anxiety in undergraduate research methods courses: Its nature and implications. Int J Res Method Educ 31(2):155–167. https://doi.org/10.1080/17437270802124616
27. Pathak V, Tripathi BD, Mishra VK (2011) Evaluation of anticipated performance index of some tree species for green belt development to mitigate traffic generated noise. Urban Forestry Urban Green 10(1):61–66. https://doi.org/10.1016/j.ufug.2010.06.008
28. Pedaste M, Mäeots M, Siiman LA, de Jong T, van Riesen SAN, Kamp ET, Manoli CC, Zacharia ZC, Tsourlidaki E (2015) Phases of inquiry-based learning: Definitions and the inquiry cycle. Educational Res Rev 14:47–61. https://doi.org/10.1016/j.edurev.2015.02.003
29. Peter S, Diekötter T, Kremer K (2019) Participant outcomes of biodiversity citizen science projects: A systematic literature review. Sustainability 11(10) Article 2780. https://doi.org/10.3390/su11102780
30. Peters K, Elands B, Buijs A (2010) Social interactions in urban parks: Stimulating social cohesion? Urban Forestry Urban Green 9(2):93–100. https://doi.org/10.1016/j.ufug.2009.11.003
31. Phillips TB, Porticella N, Constas M, Bonney R (2018) A framework for articulating and measuring individual learning outcomes from participation in citizen science. Citiz Science: Theory Pract 3(2). Article 3. https://doi.org/10.5334/cstp.126
32. Pizzolato AC, Tsuji LJS (2022) Citizen science in K–12 school-based learning settings. School Sci Math. https://doi.org/10.1111/ssm.12528

33. Prajapati SK, Tripathi BD (2008) Anticipated performance index of some tree species considered for green belt development in and around an urban area: A case study of Varanasi city, India. J Environ Manage 88(4):1343–1349. https://doi.org/10.1016/j.jenvman.2007.07.002
34. Reiss MJ (2016) Journal of Biological Education: A personal reflection on its first 50 years. J Biol Educ 50(1):3–6. https://doi.org/10.1080/00219266.2016.1140984
35. Rodenbusch SE, Hernandez PR, Simmons SL, Dolan EL (2016) Early engagement in course-based research increases graduation rates and completion of science, engineering, and mathematics degrees. CBE—Life Sci Educ 15(1). https://doi.org/10.1187/cbe.16-03-0117
36. Saha B, Haq SMA (2024) Perception of urban green space among university students in Bangladesh. PLoS ONE 19(9) Article e0311033. https://doi.org/10.1371/journal.pone.0311033
37. Semken S, Freeman CB (2008) Sense of place in the practice and assessment of place-based science teaching. Sci Educ 92(6):1042–1057. https://doi.org/10.1002/sce.20279
38. Stanfield E, Slown CD, Sedlacek Q, Worcester SE (2022) A course-based undergraduate research experience (CURE) in biology: Developing systems thinking through field experiences in restoration ecology. CBE—Life Sci Educ 21(2) Article ar20. https://doi.org/10.1187/cbe.20-12-0300
39. Weinstein N, Balmford A, DeHaan CR, Gladwell V, Bradbury RB, Amano T (2015) Seeing community for the trees: The links among contact with natural environments, community cohesion, and crime. Bioscience 65(11):1141–1153. https://doi.org/10.1093/biosci/biv151