

When Environmental Measures Disagree: Triangulating Biophilic Attributes, Pupil Experience and Academic Performance in Four Ugandan Primary Schools

Josephine Njoki Marete.¹, Wagah Mical Ongachi.², Geoffrey Kiplimo Korir³

^{1,2}Department of Art and Design, Maseno University, Kenya

³Department of Communication and Media Technology, Maseno University, Kenya

DOI: <https://doi.org/10.47772/IJRISS.2026.1026EDU0593>

Received: 15 September 2026; Accepted: 20 September 2026; Published: 28 September 2026

ABSTRACT

School-environment studies often treat observed features, pupil experience and educational outcomes as if they measured the same underlying quality. This exploratory multiple-case study examined that assumption across four government-sponsored primary schools in Mukono District, Uganda. Four raters assessed each school using a provisional 52-item presence/absence audit; 302 pupil questionnaires described perceived environment and emotional-social experience; and school records supplied separate mid-term examination summaries. Observed checklist coverage ranged from 38.5% to 78.6%, but audit reproducibility was limited (Fleiss' kappa=.21; average-measures ICC=.43, 95% CI [.16, .64]). The three evidence layers produced contrasting case profiles: School A was broadly convergent-positive, School C combined the lowest audit coverage with the second-strongest grade point, and School D combined the highest coverage with the weakest grade point. With four purposively selected schools, a provisional audit and no pupil-level linkage between surveys and examinations, these patterns may reflect contextual variation, non-equivalent measures and measurement error. They therefore generate, rather than confirm, a three-layer assessment hypothesis: physical availability, lived experience and educational outcomes should be measured separately and interpreted through their convergence before design conclusions are drawn.

Keywords: biophilic school design; environmental assessment; pupil experience; academic performance; multiple-case study; Uganda

INTRODUCTION

Biophilic school research is often organised around an intuitively compelling sequence: natural attributes are present, pupils experience the environment more positively, and improved well-being supports learning. Evidence links classroom design and children's contact with nature to comfort, restoration and learning-related outcomes (Barrett et al., 2015; Chawla, 2015; Fisher, 2024). Yet the links in this sequence represent different constructs. A feature can be physically present but inaccessible, poorly maintained or unnoticed; a positively experienced setting may reflect social and instructional conditions as well as design; and examination performance is shaped by many influences beyond the physical environment. Treating these layers as interchangeable can therefore produce confident but poorly supported design conclusions.

This distinction is especially important in under-resourced school systems. Much biophilic evidence comes from higher-income or intentionally designed settings, while African studies remain comparatively sparse and frequently concern universities or purpose-built environments (Umeora et al., 2025). Ugandan education research has instead documented pressures involving classroom provision and physical learning conditions (Asad, 2021; Ngware et al., 2016). Recent East African guidance, informed by fieldwork in Uganda, identifies respect for nature and biodiversity as a school-design principle (Feilden Foundation, 2024). However, decision-makers still need evidence that distinguishes what schools contain from how pupils experience those settings and whether either pattern aligns with educational outcomes.

Mukono District offers a bounded setting in which to examine that distinction. The district contains 187 government-sponsored primary schools and combines a rapidly urbanising municipal area with peri-urban and rural communities (Mukono District Local Government, 2019; Uganda Bureau of Statistics, 2017). The present cases comprise two peri-urban schools (A and B), one urban school (C) and one rural school (D). They were selected for contrast, not to represent their settlement categories or establish a rural-urban gradient.

An earlier publication from a broader study examined pooled pupil-level relationships between perceived biophilic dimensions and emotional-social well-being using questionnaire and qualitative evidence (Marete, 2025). It did not compare observed checklist coverage, pupil-reported experience and aggregate academic performance as separate evidence layers across school cases. The present article addresses that distinct question and does not repeat the earlier qualitative analysis.

This study therefore used a three-layer assessment framework: (1) observed attribute availability, (2) pupil-reported environmental and emotional-social experience, and (3) aggregate academic performance. It asked whether these layers produced convergent or discordant school profiles and what the patterns imply for measurement and future design research. The purpose was not to rank settlement types or test a causal chain, but to determine whether any one evidence layer could reasonably stand as a proxy for the others.

METHODS

Design and school contexts

A descriptive comparative multiple-case design was embedded within a broader convergent mixed-methods study. Four government-sponsored primary schools were purposively selected to provide environmental and settlement variation: Schools A and B were peri-urban, School C urban and School D rural. The school was the case and the unit of cross-evidence comparison. Each layer was first described independently; the resulting school profiles were then compared for convergence, partial alignment or discordance. Because there were only four purposively selected cases and context was not manipulated, no inferential test of settlement effects was attempted.

Relationship to the broader study. The environmental audit, pupil questionnaire and academic summaries originated in a study from which a related well-being article was published (Marete, 2025). That publication focused on pupil-level associations between perceived biophilic dimensions and well-being. The present secondary analysis instead treats schools as descriptive cases and examines agreement among three different evidence layers. No new participants or field data were introduced, and no previously published analysis is presented as a new primary finding.

Environmental audit

A provisional 52-item checklist recorded whether biophilic attributes were visibly present within the classrooms and associated school spaces examined. The checklist was developed for the broader study by translating established biophilic-design concepts into observable features relevant to low-resource Ugandan schools. It covered direct nature (for example, daylight, vegetation, views, air movement, water, animals and geological features), sensory and temporal qualities (natural sounds, scent, tactile variation, thermal or airflow variation, weather and seasonal change), spatial experience (prospect, refuge, movement, spatial complexity and relationships between interior and exterior areas), and natural analogues or place-based features (natural materials, colours, forms, imagery, age or patina and locally meaningful references). The instrument was designed as an attribute inventory rather than a validated psychometric scale; the broad groupings organised observation but were not scored as independent subscales. Consequently, a domain-level item count and reliability coefficient are not available. Table 1 summarises the audit's conceptual coverage.

Four raters independently recorded each item as present or absent at every school. School coverage was calculated as the number of attributes recorded present divided by 52. Binary presence did not capture quantity, condition, accessibility, duration, pupil use or the relative importance of individual features. Because the items were unequally important and the checklist was not validated as a weighted performance scale, percentages

indicate coverage only. Overall inter-rater agreement was limited: Fleiss' kappa=.21 and the average-measures ICC=.43 (95% CI [.16, .64]). Item-level and domain-level agreement statistics were not retained in the available audit summary, so the items with the lowest agreement cannot be identified reliably. This is a limitation of the instrument and of the present secondary analysis. Future applications should calculate prevalence-adjusted item agreement and domain-level coefficients, flag items with ambiguous observation boundaries, and revise those items before producing school totals. Small differences between schools were therefore not treated as diagnostic or policy-relevant.

Table 1. Conceptual coverage of the 52-item environmental audit

Audit grouping	Illustrative attributes	Scoring and reliability status
Direct nature	Daylight, plants, views, air movement, water, animals and geological features	Items contributed to the 52-item total; no separate subscale score
Sensory and temporal qualities	Natural sound, scent, touch, thermal or airflow variation, weather and seasonality	Items contributed to the 52-item total; no domain-level agreement available
Spatial experience	Prospect, refuge, movement, complexity and interior-exterior relationships	Items contributed to the 52-item total; no domain-level agreement available
Natural analogues and place	Natural materials, colours, forms, imagery, age or patina and local references	Items contributed to the 52-item total; no domain-level agreement available

For policy, this decision logic means that no single layer should trigger a high-cost intervention or a judgement about school quality. Audit scores can identify what to inspect; pupil reports can identify where provision is experienced as useful or problematic; and academic records can indicate whether broader learning support requires attention. Discordant findings should initiate verification and targeted enquiry, not cancellation of biophilic improvements or automatic expansion of feature counts.

Pupil survey

The pupil workbook contained 302 complete questionnaire records: A=23, B=101, C=106 and D=72. Pupils responded on five-point agreement scales. Four perceived-environment composites represented different aspects of biophilic experience. Nature in the Space (NIS; 10 items) addressed relatively direct contact, including perceived daylight, vegetation, views and air movement. Sensory Nature Stimuli (SNS; 10 items) addressed natural sounds, scent, touch and thermal or airflow variation; sample content included hearing birds through windows and noticing flower scent. Nature of the Space (NOS; 13 items) addressed prospect, refuge, mystery, movement, legibility and opportunities to explore. Natural Analogues (NA; 9 items) addressed indirect references to nature through colour, imagery, materials, furniture and arrangement; chair comfort was one item used in this broad study-specific composite. Emotional and Social Environment (ESE; 8 items) represented perceived comfort, calmness, safety and social experience; sample content included whether classroom lighting felt sufficient. Composite scores were arithmetic means, with higher values representing more positive perceptions.

Internal consistency in the analytical workbook was $\alpha=.715$ for NIS, $.726$ for SNS, $.704$ for NOS, $.718$ for NA and $.568$ for ESE. The four environmental composites therefore showed acceptable consistency for exploratory group description, whereas ESE's modest alpha requires caution and may attenuate or destabilise comparisons. These are study-specific perceived-environment measures, not fully validated universal scales. For each school, the descriptive perceived-environment score was the mean of the NIS, SNS, NOS and NA school means. Survey denominators differ from registers and examination summaries because they refer only to the available questionnaire records.

Academic records

Academic performance was represented by the Primary Six mid-term examination available from each school's records during the study period. Mid-term results were used because they were the most complete contemporaneous performance records accessible across the four cases and could be aligned approximately with the timing of the environmental and pupil assessments. School administrators provided aggregate grade distributions rather than pupil-identifiable marks. The distributions were converted to a mean grade point using D1=1, D2=2, D3=3, D4=4, Credit=5 and Pass/Fail=6; lower values therefore indicate stronger performance. Records covered 23 pupils in A, 110 in B, 112 in C and 61 in D.

The examinations were school-administered and were not demonstrated to use a common standardised paper, identical marking procedures or equivalent subject weighting across schools. The mean grade points must therefore be treated as contextual school summaries, not directly comparable standardised achievement scores. A single mid-term assessment may also be influenced by temporary factors, measurement error, attendance, curriculum coverage and test difficulty; it provides a less reliable indicator of sustained attainment than repeated end-of-term results or a common external examination. End-of-term or national examination data would have strengthened comparability, but they were not available in linked form for these pupils during the study window. The performance cohort was not linked pupil-by-pupil to the survey, so no individual association between experienced quality and attainment could be estimated. With N=4 schools, the analysis remained descriptive and no inferential test was used.

Ethics

Permission was obtained from the Maseno University School of Graduate Studies and the Uganda Christian University Research Ethics Committee (UG-REC-026), together with relevant education and school authorities. Parents or guardians provided informed consent, pupils provided verbal and written assent, and schools and participants were anonymised.

Results

Table 2. Comparative profiles of the four schools

School	Context	Audit coverage	Survey n	Perceived environment	ESE	Exam n	Mean grade point
A	Peri-urban	39/52 (75.0%)	23	4.05	4.28	23	2.17
B	Peri-urban	37/52 (71.2%)	101	3.72	3.87	110	2.75
C	Urban	20/52 (38.5%)	106	3.21	3.82	112	2.46
D	Rural	41/52 (78.6%)	72	3.72	3.82	61	3.91

Note. Perceived environment is the mean of NIS, SNS, NOS and NA school means. Lower mean grade point indicates stronger academic performance. Survey and examination denominators are separate.

The three evidence layers did not produce a common school ranking. School D, the rural case, had the highest checklist coverage, followed by peri-urban A and B; urban C had the lowest. Pupil-reported environmental experience was highest in A, intermediate and identical in B and D, and lowest in C. ESE was highest in A but nearly identical in B, C and D. Academic performance followed a different order: A was strongest, followed by C, B and D. These orders are descriptive and do not establish settlement effects.

Cross-case pattern matching identified one broadly convergent-positive case and two clear forms of discordance. School A combined high observed coverage, the strongest pupil-reported experience and the strongest grade point. School C was audit-achievement discordant: it had the lowest observed coverage and perceived-

environment mean but the second-strongest grade point. School D was high-audit/low-outcome discordant: it had the highest observed coverage but the weakest grade point. School B occupied an intermediate position.

Table 3 summarises these patterns as case interpretations rather than performance categories. The labels describe agreement among the available measures; they are not validated school classifications. In particular, the four school-level values cannot establish a correlation, mediation pathway or causal effect.

Table 3. Cross-evidence patterns and their analytical meaning

School	Cross-evidence pattern	What the case demonstrates
A	Convergent-positive	High audit coverage aligned with the strongest pupil experience, ESE and grade point.
B	Intermediate	Coverage similar to A did not produce equally strong pupil experience or academic performance.
C	Audit-achievement discordant	The lowest audit coverage and perception score coexisted with the second-strongest grade point.
D	High-audit/low-outcome discordant	The highest audit coverage coexisted with lower ESE and the weakest grade point.

DISCUSSION

Discordance is the central result

The central descriptive result is not that an expected relationship failed, but that three commonly conflated evidence layers produced different school profiles. Audit coverage described recorded availability, pupil scores described experienced environmental and emotional-social quality, and examination summaries described a distal outcome shaped by teaching, prior attainment, household conditions, leadership and many other factors. However, only four school profiles were compared. A pattern based on four purposively selected cases cannot establish the frequency, magnitude or statistical robustness of discordance in a wider population. It should therefore be interpreted as a hypothesis-generating observation rather than proof of a general three-layer principle.

Measurement error is also a plausible contributor. The audit's low kappa and modest ICC show that raters did not reproduce all presence/absence judgements consistently. Binary scoring may have counted a feature that was scarce, inaccessible, poorly maintained or irrelevant to pupils, while failing to distinguish a highly usable feature from a merely visible one. Some apparent disagreement between audit coverage and pupil experience may therefore arise from instrument limitations rather than genuine separation between physical provision and lived experience. The three-layer framework remains useful as a provisional assessment hypothesis because it exposes these non-equivalences, but it requires replication with a better-calibrated audit and many more schools.

What the four cases reveal

School A provides the clearest aligned pattern, but a single convergent case cannot establish a mechanism. Schools C and D are more analytically informative because they expose the limits of simple feature-count reasoning. C's stronger-than-expected academic profile indicates that low observed coverage does not imply uniformly weak school functioning. D's profile indicates that abundant recorded attributes do not guarantee stronger experience or performance. These patterns direct attention to access, condition, salience and use of features, as well as instructional and institutional conditions.

The rural, peri-urban and urban labels help describe the cases but do not explain their outcomes. Funding histories were not collected, absenteeism was not measured, and only one rural and one urban school were

included. The observed ordering should therefore generate hypotheses about urban pressure, land availability, maintenance and school organisation not be presented as a settlement gradient.

A three-layer assessment principle

For measurement, the cases motivate a provisional three-layer assessment procedure: document physical availability, investigate experienced quality and examine relevant outcomes separately, then interpret where the layers converge or diverge. The procedure does not assume that academic outcomes should immediately respond to environmental provision. Instead, discordance prompts a diagnostic question about access, condition, use, measurement quality or non-environmental influences. Future audits should define observation boundaries, distinguish presence from quantity, condition, accessibility and use, pilot culturally relevant items, and employ rater calibration, paired photographs and instrumental measures. Pupil voice remains a distinct rather than substitute source of evidence.

Table 4. Practical interpretation of three-layer evidence patterns

Observed pattern	Interpretation	Recommended response
Audit high; pupil experience poor	Features may be inaccessible, poorly maintained, unevenly distributed or not salient; audit error is also possible.	Conduct pupil walkthroughs and seating maps; inspect condition, access and use; recalibrate disputed audit items before adding more features.
Audit and pupil experience high; performance low	The environment may support comfort without overcoming instructional, attendance, assessment or socioeconomic constraints.	Retain supportive features; examine teaching, attendance, test comparability and learning resources; use repeated linked outcomes before attributing performance.
Audit low; pupil experience high	A small number of valued features, social climate or unmeasured qualities may matter more than feature count.	Protect the features pupils value; document how they are used; avoid assuming that a low checklist total means poor lived quality.
Audit low; pupil experience and performance low	Environmental deficits may coexist with wider institutional constraints.	Prioritise safe daylight, ventilation, views, shade and comfort alongside instructional and resource support; monitor all three layers after improvement.
All three layers high	The profile is encouraging but does not by itself demonstrate causation.	Maintain provision, document implementation and test whether the pattern persists longitudinally and across comparable schools.

Strengths and limitations

A strength of the study is its explicit separation and triangulation of observational, experiential and outcome evidence, together with transparent reporting of rater disagreement and non-equivalent denominators. Its limits are equally important: four purposively selected schools, one time point, a provisional binary audit with low agreement, aggregate rather than pupil-linked academic data, and no measured instructional, socioeconomic, attendance or funding-history covariates. The examination scores came from one non-standardised mid-term assessment and may not be comparable across schools. The design supports analytic insight into possible measurement discordance, not a robust estimate of its prevalence, causal inference or settlement-level generalisation.

Audit development should begin with an item manual containing operational definitions, observation boundaries, photographic examples and explicit rules for borderline cases. Raters should complete shared training and an independent pilot in schools outside the main sample. Item-level agreement and prevalence

should then be reviewed; ambiguous or near-universal items should be revised or removed. Presence, quantity, condition, accessibility and observed use should be recorded separately where feasible. Calibration should continue until acceptable item- and domain-level agreement is demonstrated, after which a confirmatory reliability study can establish whether school totals are sufficiently reproducible for comparative use.

For planning, a simple pupil-level correlation of $r=.20$ requires approximately 194 independent observations for 80% power at $\alpha=.05$. Clustering substantially increases that requirement: with about 20 pupils per school and an intraclass correlation near .05, the design effect is about 1.95, suggesting roughly 380 linked pupils before allowance for attrition or additional covariates. A practical target would be at least 20-30 schools and approximately 400-600 pupils, followed by simulation-based power analysis using plausible school-level variance and the intended longitudinal model. More schools, rather than many pupils from only a few schools, are essential for separating institutional from pupil-level effects.

Future studies can create pupil-level linkage by assigning each consenting pupil a confidential study code used on the questionnaire and on an examination-extraction form maintained by an authorised school liaison. The linkage key should be encrypted, stored separately and destroyed after verified matching; analysis files should contain only anonymised codes. Repeated end-of-term scores, or a common externally moderated assessment, should be collected together with prior attainment, attendance and relevant background variables. This would allow multilevel models in which pupils are nested within schools and measured across time.

CONCLUSION

Across four Ugandan primary schools, observed biophilic coverage, pupil experience and mid-term academic summaries did not produce the same ordering. Because the audit had limited inter-rater reliability, examinations were not standardised and only four unlinked school cases were available, the pattern does not confirm a general law of discordance. It offers a testable assessment hypothesis: availability, lived quality and educational outcomes are distinct layers, and their apparent convergence should be verified rather than assumed. For practice, disagreement is a prompt to inspect access, condition, pupil use, measurement quality and wider instructional constraints. For research, the next step is a calibrated audit, pupil-linked repeated achievement data and a sufficiently large multisite longitudinal design. Until that evidence is available, audit scores should formulate design questions, pupil accounts should guide inclusive improvement, and academic summaries should not be treated as direct tests of environmental quality.

REFERENCES

1. Asad, K. W. (2021). Primary and secondary education in Uganda: Challenges and prospects. *Interdisciplinary Journal of Education*, 4(1). <https://doi.org/10.53449/ije.v4i1.152>
2. Barrett, P., Davies, F., Zhang, Y., & Barrett, L. (2015). The impact of classroom design on pupils' learning: Final results of a holistic, multi-level analysis. *Building and Environment*, 89, 118–133. <https://doi.org/10.1016/j.buildenv.2015.02.013>
3. Chawla, L. (2015). Benefits of nature contact for children. *Journal of Planning Literature*, 30(4), 433–452. <https://doi.org/10.1177/0885412215595441>
4. Feilden Foundation. (2024). *East African Schools Design Guide*. Feilden Foundation in partnership with Enabel.
5. Fisher, K. (2024). The biophilic school: A critical synthesis of evidence-based systematic literature reviews. *Architecture*, 4(3), 457–478. <https://doi.org/10.3390/architecture4030025>
6. Marete, J. N. (2025). The impact of biophilic interior design on pupil well-being in Ugandan primary schools: A convergent mixed-methods study. *International Journal of Research and Innovation in Social Science*, 9(10), 7961–7967. <https://doi.org/10.47772/IJRISS.2025.910000653>
7. Mukono District Local Government. (2019). *State of the district report, FY 2018/19*. <https://mukono.go.ug/publications/state-district-report-fy-1819-chairperson-lcv>
8. Ngware, M., Hungi, N., Mahuro, G., Mutisya, M., & Abuya, B. (2016). The quality of education in Uganda: A case of Iganga and Mayuge Districts. *African Population and Health Research Center*. <http://dx.doi.org/10.13140/RG.2.2.18273.92004>



9. Uganda Bureau of Statistics. (2017). National Population and Housing Census 2014: Area specific profile series—Mukono District. Uganda Bureau of Statistics.
10. Umeora, C. O., Onwuzuligbo, C. C., & Ononuju, F. I. (2025). The role of biophilic campus design in enhancing university environments in southeast Nigeria. *Open Journal of Environmental Research*, 6(1), 18–27. <https://doi.org/10.52417/ojer.v6i1.858>