

Article

Short-Term Knowledge Gains and Regional Heterogeneity in a STEM-Based Indoor Air Quality Education Intervention for Sustainability Across Asian Regions

Wen-Jing Deng ^{1,*} , Jiayue Sun ¹, Wingkei Ho ¹  and John Chi-Kin Lee ^{2,3} 

¹ Department of Science and Environmental Studies, The Education University of Hong Kong, Tai Po, N.T., Hong Kong, China; sunj@eduhk.hk (J.S.); keithho@eduhk.hk (W.H.)

² Academy for Applied Policy Studies and Education Futures, The Education University of Hong Kong, Tai Po, N.T., Hong Kong, China; jcklee@eduhk.hk

³ Centre for Environment and Sustainable Development, The Education University of Hong Kong, Tai Po, N.T., Hong Kong, China

* Correspondence: wdeng@eduhk.hk; Tel.: +852-2948-8288

Abstract

Indoor air quality (IAQ) education can support sustainability education by linking environmental health literacy, school environmental quality, and daily exposure mitigation behaviors. However, comparative evidence on short-term IAQ knowledge change across diverse educational contexts remains limited. In this study, we conducted a pre-test/post-test program evaluation examining immediate IAQ knowledge change among 1408 students (Grades 5–10) in Sri Lanka ($n = 395$), Nepal ($n = 300$), Malaysia ($n = 314$), Indonesia ($n = 200$), and Guangxi, China ($n = 199$). The primary outcome was knowledge score gain (post minus pre, 0–100 scale). We used paired t -tests with 95% confidence intervals (CI) and Cohen's d_z , ANOVA and Kruskal–Wallis tests for regional heterogeneity, BH-adjusted Welch pairwise tests, OLS gain models, and sensitivity analyses for post-test zero records. The overall mean score increased from 47.37 to 56.62 (gain +9.25, 95% CI 7.58 to 10.92; $p = 2.30 \times 10^{-26}$; $d_z = 0.289$). Gains differed substantially by region: Nepal +22.25, Indonesia +15.93, Sri Lanka +5.39, Malaysia +5.36, and Guangxi (China) −3.27 ($p = 0.081$). Region-level heterogeneity was significant (ANOVA $F = 26.55$, $p = 2.98 \times 10^{-21}$, $\eta^2 = 0.070$; Kruskal–Wallis $H = 84.95$, $p = 1.56 \times 10^{-17}$). Baseline score was the strongest predictor of gain (beta = −0.686, $p < 0.001$; model $R^2 = 0.256$). Excluding post = 0 with pre > 0 records increased the overall gain to +12.93, while the Guangxi result remained negative. The findings can serve as short-term program-evaluation evidence suggesting that IAQ sustainability education requires baseline-informed and context-sensitive curriculum adaptation, while future studies should include comparison groups, delayed post-tests, behavioral outcomes, and stronger implementation-fidelity measures.

Keywords: indoor air quality education; sustainability education; STEM intervention; short-term knowledge gain; regional heterogeneity; baseline readiness; program evaluation



Academic Editor: António José Candeias Curado

Received: 6 June 2026

Revised: 7 July 2026

Accepted: 10 July 2026

Published: 14 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Indoor air quality (IAQ) is increasingly recognized as an important issue for children's health, well-being, and learning, yet it remains comparatively underemphasized in environmental education [1,2]. Students spend much of their time indoors, especially in homes and schools, where they may be exposed to particulate matter, carbon monoxide, volatile organic compounds, biological contaminants, smoke from solid fuels, and

poorly ventilated indoor microenvironments [3,4]. The consequences of poor IAQ extend beyond immediate discomfort. They are linked to respiratory disease, cardiovascular burden, impaired well-being, and reduced quality of life. Evidence from recent reviews further suggests that indoor environmental conditions in schools are closely related to pupils' health, attendance, and academic performance [1,5,6]. Therefore, IAQ is not only a technical environmental-health issue but also an educational issue that concerns awareness, interpretation, preventive behavior, and informed decision-making.

In public discourse, however, air pollution education often privileges outdoor pollution while indoor exposure remains less visible. This imbalance is reinforced by the fact that air pollution is often misperceived by the public and that people's perception of air quality does not always correspond to actual exposure or risk [7]. Students may therefore learn about smog, traffic emissions, or climate change without receiving equivalent instruction about cooking smoke, inadequate ventilation, off-gassing from materials, indoor combustion, classroom crowding, or behavioral strategies that reduce exposure in daily life. Yet indoor environments are among the spaces where children and adolescents can most readily apply environmental knowledge. If environmental education seeks to connect knowledge with practical behavior, IAQ is a particularly suitable domain in which to study that connection [8,9].

Research on environmental education has demonstrated that targeted interventions can improve learners' knowledge, attitudes, and perceptions, particularly when instruction is inquiry-based, contextually relevant, and connected to everyday life [10–12]. Recent conceptual work has also emphasized that environmental and sustainability education should move beyond simple information transmission and engage learners in systems literacy, situated reasoning, and the interpretation of complex environmental problems [13]. Pre-test and post-test designs have also been widely used to assess short-term knowledge gains in environmental education, especially through multiple-choice instruments that allow conceptual understanding to be measured consistently across groups [11,14]. At the same time, intervention effects are rarely uniform. They depend on learner readiness, age, curriculum alignment, cultural expectations, pedagogical design, and the extent to which students perceive the content as relevant to their own lives [10].

For IAQ education specifically, comparative evidence remains limited. Our previous study evaluated a STEM-based IAQ program in Nepal, Malaysia, Indonesia, and Guangxi, China, using data from 1027 students in 11 schools across Grades 1–9 [15]. That study examined multidimensional self-reported outcomes, including knowledge, behaviour, attitude, and expectation, and provided broad evidence on STEM-based IAQ education for sustainable development. The present study extends this work by adding Sri Lanka, using a larger five-region sample of 1408 students from 13 schools across Grades 5–10, and focusing on immediate objective IAQ knowledge gain measured using multiple-choice items. The two studies therefore address different questions: the previous study assessed broader perceived learning outcomes, whereas the present study examines how objective short-term knowledge gain varied across contexts and learner groups.

The five study settings—Sri Lanka, Nepal, Malaysia, Indonesia, and Guangxi, China—differ in household energy profiles, school infrastructure, urbanization patterns, environmental discourse, and likely baseline familiarity with IAQ concepts [6,16–18]. Household energy use and ventilation are especially relevant to IAQ education in Nepal and Sri Lanka [19,20], while classroom indoor environmental quality and environmental education have received increasing attention in Malaysia [21,22]. In Guangxi, China, prior research on environmental citizenship and environmental education suggests a different baseline context for environmental awareness [23,24]. Recent reviews of school IAQ further indicate that educational and technical interventions should be interpreted in relation to local

school conditions [25]. These contextual differences make cross-region variation central to evaluating how a shared IAQ curriculum may perform when implemented across diverse educational settings.

Accordingly, the study evaluates short-term pre–post knowledge gain after a standardized STEM-based IAQ education activity and examines whether gains differed by region, grade, sex, and baseline knowledge. The study does not aim to claim causal effectiveness because it lacks an untreated comparison group, but it provides evidence on patterns of immediate knowledge change, heterogeneity across learner groups, and the sensitivity of interpretation to baseline score and post-test zero records. The analysis addresses three questions: what immediate IAQ knowledge change was observed after the activity; how gains varied across region, grade, and sex; and how baseline score and post-test zero records affected interpretation of the observed pattern. These findings inform the design of scalable IAQ sustainability education by clarifying when a common curriculum may require differentiated examples, difficulty levels, and local implementation support.

2. Methods

2.1. Study Design and Participants

The study used a single-group pre-test and post-test design. Students completed a baseline IAQ knowledge assessment before participating in the education program and an immediate post-test after the intervention. Program delivery and evaluation took place between June 2024 and February 2025 across five regions: Sri Lanka, Nepal, Malaysia, Indonesia, and Guangxi, China. The initial dataset contained 1472 student records. Because the gain-score analysis required matched pre-test and post-test data, 64 records with only a usable pre-test or only a usable post-test were excluded. The final analytic sample included 1408 paired observations: Sri Lanka ($n = 395$), Nepal ($n = 300$), Malaysia ($n = 314$), Indonesia ($n = 200$), and Guangxi, China ($n = 199$). The sample included 775 girls and 633 boys from Grades 5 to 10; detailed grade and sex distributions are shown in Table 1.

Table 1. Participant distribution by region, grade, and sex.

Characteristic	Sri Lanka	Nepal	Malaysia	Indonesia	Guangxi, China	Total
Grade 5	0	0	57	0	0	57
Grade 6	10	0	63	0	0	73
Grade 7	125	76	147	11	199	558
Grade 8	178	41	47	38	0	304
Grade 9	74	183	0	151	0	408
Grade 10	8	0	0	0	0	8
Female	242	136	162	130	105	775
Male	153	164	152	70	94	633

Participants were recruited from 13 primary and secondary schools across the five study regions. School participation depended on registration, local implementation capacity, and available staff and resources, so the study should be understood as a feasible implementation sample rather than a random sample of students in each setting. The study was conducted under university ethics approval through The Education University of Hong Kong (No. 2024/25-0233; 17 December 2024), together with school-level permission and parental consent for minor participants.

2.2. Educational Intervention

The intervention was a one-day, six-hour STEM-based environmental education activity focused on IAQ. The session consisted of four parts: classroom instruction on IAQ

concepts and health effects (2 h), practical activities using IAQ monitoring equipment to collect real-time indoor environmental data (2 h), guided data analysis and interpretation (1 h), and group reflection and feedback (1 h). Its main purpose was to strengthen students' understanding of indoor pollution sources, health effects, measurement, and practical mitigation strategies while connecting IAQ to environmental health literacy and sustainability-oriented decision-making in schools and homes. The design was aligned with experiential and situated approaches to environmental learning, which emphasize active engagement, inquiry, and connection to learners' everyday ecological contexts [8,26].

To maintain cross-site comparability, the research team coordinated the educational framework with local implementing teachers so that all regions addressed the same core IAQ concepts and practical monitoring tasks. Instructional materials included presentations, worksheets, posters, videos, and hands-on IAQ monitoring instruments. Local examples and language were adapted where needed, but the core learning sequence was kept consistent. The same core intervention was used across grade levels; differences by age were handled through explanation depth, examples, and task support rather than through separate curricula. No participants had previously taken part in the program.

2.3. Measures and Procedure

Knowledge outcomes were assessed using short multiple-choice instruments administered immediately before and after the activity. The tests were designed to capture students' understanding of basic IAQ concepts, pollutant sources, health effects, measurement, and mitigation practices. All regional versions used three pre-test knowledge items and four post-test knowledge items. The items were coded 1 for a correct answer and 0 for an incorrect answer, and total scores were converted to a 0–100 scale to support comparison across sites and item counts. The assessment contained short test items, so it was not suitable for stable estimation of internal consistency reliability. The resulting scores could be interpreted as brief indicators of immediate IAQ knowledge, rather than as comprehensive measures of students' overall IAQ competence. The pre- and post-tests assessed the same content domain but were not identical item sets. Example item stems and answer-option content are provided in Supplementary Materials Table S1.

The same core grade-level version was used across Grades 5–10 within each site. A common item framework was used across implementation sites, and the questions were translated or converted into the local instructional language for regional delivery while preserving the same IAQ concepts, item intent, and scoring logic. Local examples and wording were adjusted where needed to support comprehension, but the tests were kept within the same content domain across regions and grade levels. The instrument was developed in collaboration with environmental health and education specialists and piloted in Hong Kong between 2016 and 2019 to refine clarity, difficulty, and usability. To reduce order effects and routine response patterns, the order of questions and answer options was randomized in paper and online administration formats where feasible. Trained researchers supervised the assessment process, explained completion requirements, and ensured that students worked independently.

2.4. Data Preparation and Outcome Variables

The primary analytic sample consisted of students with complete and usable paired pre-test and post-test knowledge scores. For the students who appeared only in the pre-test or only in the post-test, the records could not contribute to paired gain analysis and accounted for the 64 records excluded from the initial dataset. The details of data cleaning are described in Supplementary Materials Section S2.

The primary outcome was gain score, defined as post-test score minus pre-test score on the 0–100 scale. Positive values indicate improvement, zero indicates no change, and negative values indicate decline. In addition to analyzing mean gain, the study classified students into improved, unchanged, and declined groups to describe the distribution of outcomes beyond average score shifts alone.

Before inferential analysis, the dataset was screened for missingness, arithmetic consistency of score fields, and unusual score patterns. The final analytic file contained 1408 observations, no missing values, and complete agreement between the recorded change column and the calculated post-minus-pre difference. Records with post-test scores of zero despite non-zero pre-test scores were retained in the primary analysis because they could represent true scores, nonresponse, recording problems, or administration issues. Their distribution by anonymized implementation source and the sensitivity analysis excluding these records are reported in Supplementary Materials Tables S2–S4.

2.5. Statistical Analysis

The analysis proceeded from overall pre–post change to subgroup heterogeneity, adjusted modelling, and robustness assessment. Descriptive statistics were calculated for the full sample and for major subgroups. Paired *t*-tests were used to evaluate pre–post differences overall and within region, grade, and sex categories, with 95% confidence intervals and Cohen’s *d* reported to describe the magnitude of within-student change. For paired comparisons, Cohen’s *d* was calculated as the mean gain divided by the standard deviation of the within-student gain scores.

Between-region heterogeneity in gain was examined using both one-way analysis of variance and the Kruskal–Wallis test. Pairwise regional contrasts were then estimated using two-sided Welch *t*-tests with Benjamini–Hochberg adjustment to control the false discovery rate across multiple comparisons, and Cohen’s *d* was reported for the size of between-region differences. Because grade distribution differed strongly by region, grade-specific results were interpreted descriptively rather than as independent developmental effects.

To assess whether regional patterns persisted after adjustment for student characteristics, an ordinary least squares model was fitted with gain score as the dependent variable and baseline score, grade, sex, and region indicators as predictors. Baseline score was included to account for baseline-related headroom and ceiling effects. A robustness model with cluster-robust standard errors using the available district/school-source grouping variable was also estimated to address potential dependence among students from the same implementation setting. Additional implementation-source summaries and graphical robustness checks are provided in the Supplementary Materials.

3. Results

3.1. Overall Pre–Post Change

Across all 1408 students, the mean pre-test score was 47.37, and the mean post-test score was 56.62. The mean gain was 9.25 points (95% CI 7.58 to 10.92). The paired *t*-test indicated a significant increase in score after the intervention ($t = 10.84, p = 2.30 \times 10^{-26}$), with a small-to-moderate within-student effect size (Cohen’s *d* = 0.289).

Outcome composition showed that 53.48% of students improved, 9.73% showed no change, and 36.79% declined. These proportions are shown in Figure 1.

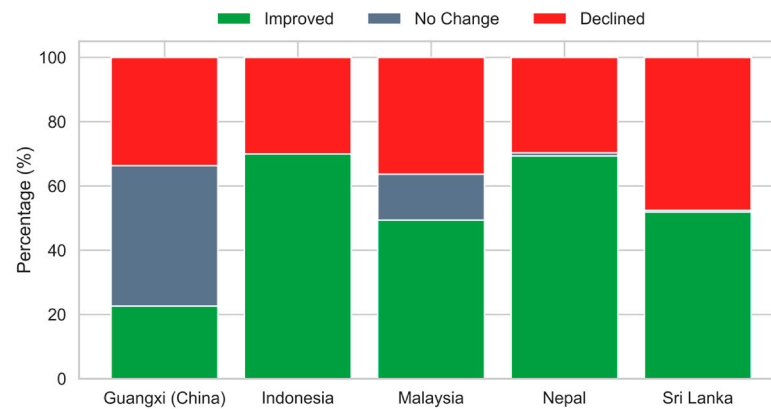


Figure 1. Regional composition of improved, unchanged, and declined outcomes ($n = 1408$).

3.2. Regional Differences in Gain

Gain differed markedly across the five regions. Figure 2 presents the pre-test and post-test mean scores for each region. Nepal showed the largest increase, with scores rising from 35.58 to 57.83 and a mean gain of 22.25 points (95% CI 18.58 to 25.93, $p = 4.78 \times 10^{-27}$, $dz = 0.688$). Indonesia also showed a substantial increase, from 44.58 to 60.50, corresponding to a mean gain of 15.93 points (95% CI 11.61 to 20.24, $p = 7.98 \times 10^{-12}$, $dz = 0.514$). Sri Lanka and Malaysia showed smaller but statistically significant gains of 5.39 and 5.36 points, respectively. Guangxi (China) differed from the other regions, with a pre-test mean of 83.54, a post-test mean of 80.28, and a non-significant mean change of -3.27 points (95% CI -6.94 to 0.41 , $p = 0.081$).

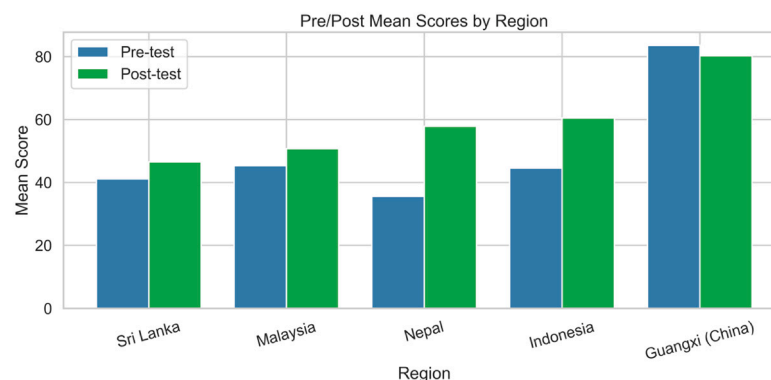


Figure 2. Pre- and post-test mean scores across five regions (Sri Lanka $n = 395$; Nepal $n = 300$; Malaysia $n = 314$; Indonesia $n = 200$; Guangxi, China $n = 199$).

The heterogeneity of regional gain was supported by both parametric and non-parametric tests. One-way ANOVA yielded $F = 26.55$, $p = 2.98 \times 10^{-21}$, with eta squared = 0.070, while the Kruskal–Wallis test yielded $H = 84.95$, $p = 1.56 \times 10^{-17}$. The regional pre-test and post-test mean scores with 95% confidence intervals are shown in Figure 3.

Figure 4 further indicates that the regional pattern was not limited to differences in mean values, but also involved differing within-region gain distributions. Pairwise comparisons provided additional statistical detail. After Benjamini–Hochberg adjustment, Guangxi (China) differed significantly from Nepal, Indonesia, Sri Lanka, and Malaysia. The largest adjusted contrast was between Guangxi (China) and Nepal, with a mean difference of -25.52 points and a large effect size ($d = -0.849$). Nepal also differed significantly from both Sri Lanka and Malaysia, whereas Malaysia and Sri Lanka did not differ significantly from one another after correction.

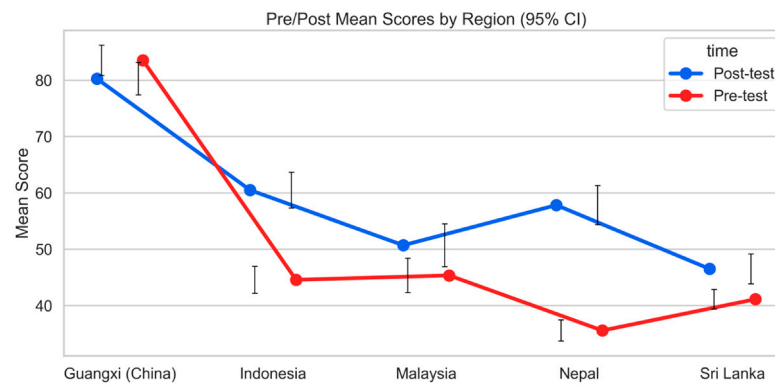


Figure 3. Regional pre- and post-test mean scores with 95% confidence intervals.

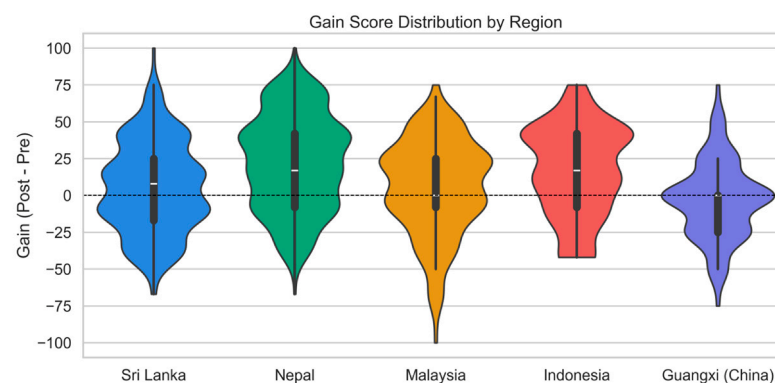


Figure 4. Distribution of gain scores by region. Regional sample sizes are Sri Lanka $n = 395$, Nepal $n = 300$, Malaysia $n = 314$, Indonesia $n = 200$, and Guangxi, China $n = 199$.

3.3. Grade and Sex Subgroup Patterns

When analysed by sex, both female and male students showed significant pre–post gains. Female students increased by 9.75 points on average ($p = 1.50 \times 10^{-17}$, $d_z = 0.314$), while male students increased by 8.64 points ($p = 1.13 \times 10^{-10}$, $d_z = 0.261$). Figure 5 shows the sex-specific distributions, and Figure 6 shows regional variation in mean gain by sex.

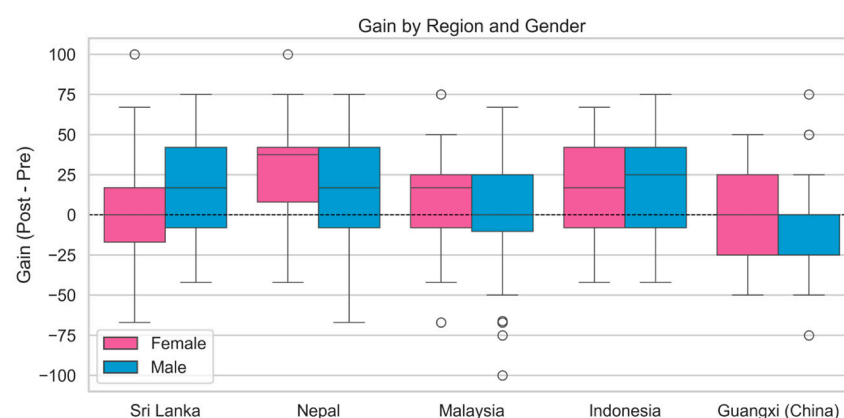


Figure 5. Gain scores by region and gender (female $n = 775$; male $n = 633$). Boxes show the interquartile range with median lines, whiskers show 1.5 times the interquartile range, dots indicate outliers, and the dashed line marks zero gain.

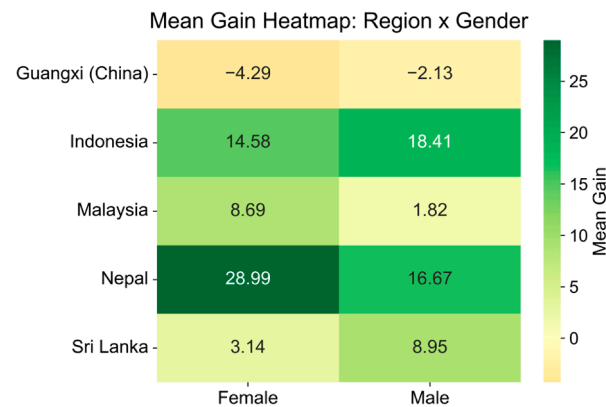


Figure 6. Heatmap of mean gain scores by region and gender. Cell values represent mean gain scores.

Grade-specific results were more heterogeneous. As shown in Figure 7, Grade 9 showed the largest gain (+22.41, $p = 2.02 \times 10^{-40}$, $dz = 0.738$), followed by Grade 6 (+17.73, $p = 3.74 \times 10^{-7}$, $dz = 0.655$) and Grade 5 (+14.91, $p = 1.67 \times 10^{-4}$, $dz = 0.535$). Grade 8 also showed a significant increase (+6.68, $p = 9.11 \times 10^{-4}$, $dz = 0.192$). By contrast, Grade 7 showed little overall change (−0.72, $p = 0.552$), and the estimate for Grade 10 was imprecise because the subgroup was small ($n = 8$).

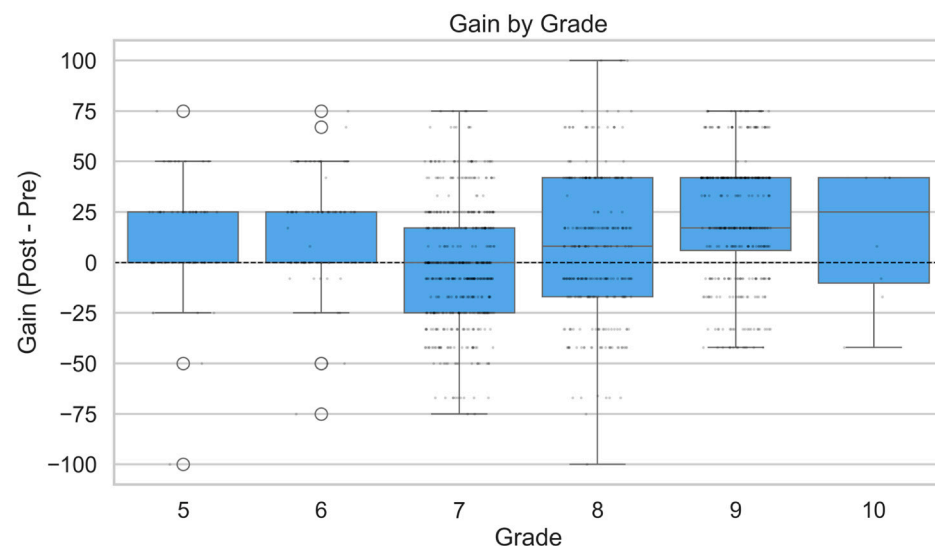


Figure 7. Gain scores across grade levels. Grade-level sample sizes are Grade 5 $n = 57$, Grade 6 $n = 73$, Grade 7 $n = 558$, Grade 8 $n = 304$, Grade 9 $n = 408$, and Grade 10 $n = 8$. Boxes indicate the interquartile range, horizontal lines within boxes indicate medians, whiskers indicate 1.5 times the interquartile range, dots represent individual student gain scores, and the dashed horizontal line indicates no change in score.

3.4. Baseline Score and Adjusted Model

In the ordinary least squares (OLS) model, baseline pre-test score was the strongest predictor of gain (beta = −0.686, $p < 0.001$). The model explained 25.6% of the variance in gain ($R^2 = 0.256$). The negative coefficient indicates that students with higher starting scores tended to show smaller subsequent gains, consistent with limited headroom on the bounded 0–100 score scale. In Table 2, after adjustment for baseline score, grade, sex, and region, some regional differences remained evident in the model output, whereas grade and sex were not significant predictors in the adjusted model. Implementation-source cluster-robust estimates and leave-one-source-out checks in the Supplementary Materials further show that the baseline-score association was stable, while regional coefficients were

less precise under clustered uncertainty. The relationship between baseline score and gain is shown in Figure 8.

Table 2. OLS gain model with conventional and cluster-robust standard errors ($N = 1408$; $R^2 = 0.256$; cluster-robust SE based on 11 available district/school-source groups).

Predictor	Coef.	SE	<i>p</i>	Cluster Coef.	Cluster SE	Cluster <i>p</i>
Intercept	56.283	8.285	<0.001	56.283	34.583	0.135
Indonesia (ref: Guangxi, China)	−7.001	3.434	0.042	−7.001	16.113	0.673
Malaysia (ref: Guangxi, China)	−17.751	2.942	<0.001	−17.751	14.881	0.260
Nepal (ref: Guangxi, China)	−6.593	3.244	0.042	−6.593	11.301	0.573
Sri Lanka (ref: Guangxi, China)	−20.221	2.926	<0.001	−20.221	8.238	0.034
Baseline pre-test score	−0.686	0.037	<0.001	−0.686	0.106	<0.001
Grade	−0.476	0.976	0.626	−0.476	5.338	0.931
Female (ref: male)	2.114	1.501	0.159	2.114	3.377	0.545

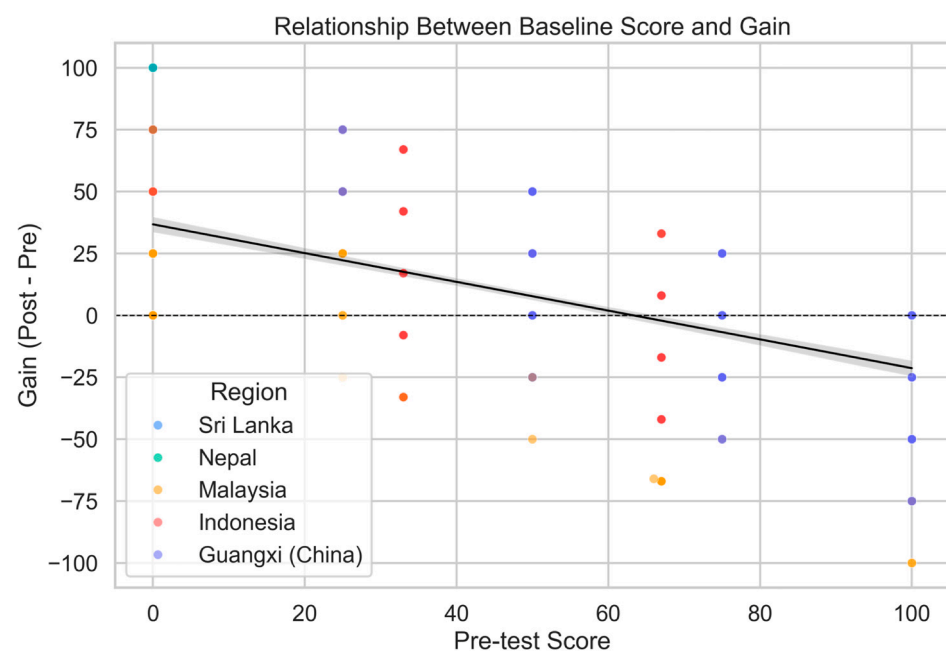


Figure 8. Association between baseline (pre-test) score and gain across the analytic sample ($N = 1408$). Apparent intermediate color shades result from the overlap of semi-transparent points and do not indicate additional groups or variables. The solid black line represents the fitted linear regression, the grey shaded band shows the 95% confidence interval, and the dashed horizontal line marks zero gain.

3.5. Sensitivity Analysis

The cleaned analytic dataset contained 1408 observations and no missing values. The recorded change column matched the calculated difference between post-test and pre-test scores for all observations. At the same time, 8.17% of observations had post-test scores of zero, 6.68% had pre-test scores of zero, and 8.17% had pre-test scores at or above 90. The 98 records with post = 0 and pre > 0 were unevenly distributed across regions: Sri Lanka 39, Malaysia 35, Nepal 23, Indonesia 1, and Guangxi (China) 0. Supplementary Materials

Figure S4 reports the same issue by anonymized implementation source and includes both the number of records and the source-specific analytic sample size.

Sensitivity analysis excluded the 98 records in which the post-test score was zero while the pre-test score was above zero, reducing the sample from 1408 to 1310 observations. These records were further examined due to the possibility that they represented true zero scores, nonresponse, coding errors, or administration-related issues. Nevertheless, the final analytic dataset lacked sufficient information to differentiate these scenarios at the individual-record level. Accordingly, the sensitivity analysis was performed to evaluate the robustness of the primary results, rather than to determine the validity or invalidity of these records. Under this specification, the overall mean gain increased from 9.25 to 12.93 points, and Cohen's d_z increased from 0.289 to 0.434 ($p = 4.79 \times 10^{-51}$). This suggests that the primary estimate is conservative if these zero post-test records reflect nonresponse or coding problems. However, the records were retained in the main analysis because their source could not be determined definitively from the final analytic file.

Regional results under sensitivity analysis showed larger gains in several settings, as shown in Table 3, with increases to 26.96 in Nepal, 16.17 in Indonesia, 12.03 in Malaysia, and 9.98 in Sri Lanka. In contrast, Guangxi (China) remained negative (−3.27). Since Guangxi contained no records with post-test scores of zero and pre-test scores above zero, its negative mean change cannot be attributed to this potential data-quality issue.

Table 3. Distribution of post = 0 with pre > 0 records by region.

Region	Analytic n	Post = 0 & Pre > 0 Records	Interpretive Note
Sri Lanka	395	39	May reflect nonresponse, recording problems, or true zero post-test scores
Nepal	300	23	Concentrated in a subset of implementation settings
Malaysia	314	35	Concentrated in 2 school and Grade 7–8 records
Indonesia	200	1	Minimal influence on regional estimate
Guangxi, China	199	0	Negative mean change is not explained by this data-quality issue
Total	1408	98	Excluded only in sensitivity analysis; retained in primary analysis

The supplementary cluster-robust sensitivity model produced a nearly identical baseline-score coefficient after excluding these records (from −0.686 to −0.679, both $p < 0.001$), further supporting the robustness of the baseline-readiness interpretation.

4. Discussion

4.1. Principal Findings

This study examined immediate IAQ knowledge change following a standardized STEM-based sustainability education program implemented across five Asian regions. The findings show an overall pre–post increase in IAQ-related knowledge, but this average improvement masked substantial variation across regions and learner groups. Baseline pre-test score was the strongest predictor of gain, indicating that students' starting knowledge was closely related to the amount of measurable short-term improvement. These findings are consistent with synthesis evidence showing that environmental education often pro-

duces positive learning outcomes, while effect sizes vary across studies, populations, and implementation contexts [27,28].

The results therefore support two linked interpretations. First, a shared IAQ curriculum can generate short-term knowledge improvement across diverse educational settings. Second, the magnitude and direction of observed gain depend strongly on baseline readiness and local implementation context. This distinction is important for sustainability education because scalable programs often rely on common curricular frameworks, yet their educational effects may differ when learners enter with different prior knowledge, school conditions, and exposure-related experiences.

4.2. Regional Heterogeneity and Baseline Readiness

Regional heterogeneity was one of the most important findings. Nepal and Indonesia showed large gains, Sri Lanka and Malaysia showed more modest positive gains, and Guangxi, China did not show significant positive improvement under the current intervention format. These differences suggest that standardization alone is insufficient for interpreting or scaling IAQ education. A curriculum that is suitable for one setting may be too elementary, too familiar, or insufficiently connected to students' local experiences in another. This interpretation is consistent with recent evidence that scalable environmental education depends on how common program frameworks are adapted to local socio-cultural and implementation conditions across regions and school systems [29].

Baseline readiness helps explain part of this variation. Students with higher pre-test scores generally showed smaller gains, which is expected when outcomes are measured on a bounded 0–100 scale and when the intervention is introductory in scope. Learners with lower baseline knowledge had more room for measurable improvement, whereas learners who were already familiar with IAQ concepts may have required more advanced content, more challenging inquiry tasks, or more locally specific applications.

The Guangxi, China results are consistent with this interpretation but also require caution. This subgroup had the highest mean pre-test score and showed a small negative mean change after the intervention. This pattern should not be interpreted as evidence that the program was ineffective in that setting. It more likely indicates that the common introductory curriculum was less well calibrated for a group with higher baseline knowledge. The absence of post-test-zero records in Guangxi, China also shows that this pattern was not driven by the data-quality issue examined in the sensitivity analysis.

4.3. Gender, Grade, and Subgroup Interpretation

Sex was not a major source of variation in the present dataset. Both female and male students showed significant pre–post gains, and the difference between them was modest after adjustment. Grade-level results require more caution. Grades were unevenly distributed across regions: all Guangxi students were in Grade 7, Grade 5 students came from Malaysia only, and Grade 10 included only a small number of students from Sri Lanka. As a result, pooled grade means partly reflect regional composition and should not be interpreted as independent developmental effects.

These subgroup findings highlight the importance of interpreting educational outcomes in relation to sample structure. Apparent differences by grade or region may reflect overlapping differences in curriculum exposure, school context, prior familiarity, and implementation conditions. For multi-site sustainability education evaluations, descriptive subgroup comparisons should therefore be supported by adjusted models and sensitivity analyses rather than interpreted from pooled means alone.

4.4. Implications for Scalable IAQ Sustainability Education

The findings suggest that IAQ education is well suited to sustainability education because it connects scientific understanding, health literacy, and everyday environmental decision-making. Students can relate IAQ concepts to familiar indoor settings such as classrooms and homes, where ventilation, indoor pollution sources, monitoring evidence, and exposure reduction can be made concrete. This relevance may support short-term learning, but the present results also show that relevance alone is not enough. Educational challenge, prior knowledge, and contextual fit remain important [7]. Existing school IAQ intervention research has shown the value of technical measures such as ventilation, air cleaning, source control, positive-pressure fresh-air systems, and monitoring, but these measures also require users to understand why, when, and how IAQ actions should be taken [25,30]. The present findings extend that literature by showing that a STEM-based IAQ curriculum can support immediate knowledge gain while also revealing that the same curriculum may not work evenly across learner groups with different baseline knowledge.

For scalable implementation, the results support a baseline-informed approach. A common curriculum can provide a stable framework for program delivery and evaluation, but its examples, difficulty level, and inquiry tasks may need adjustment across settings. In lower-readiness settings, instruction may need to emphasize basic IAQ concepts, pollution sources, ventilation, and health links. In higher-readiness settings, the same program may need to move more quickly toward data interpretation, mitigation strategies, and local action planning. This approach is consistent with tiering principles in environmental assessment, where information is organized so that subsequent decisions focus on issues most relevant to the current level of action [31].

A context-sensitive approach is also important because IAQ education is shaped by socio-cultural and regional conditions. Prior work has emphasized the need to consider socio-cultural dynamics when developing culturally relevant educational strategies and policies [32]. Future IAQ education programs should therefore examine how local indoor exposure scenarios, language, household practices, school infrastructure, and culturally specific examples affect both immediate learning and longer-term sustainability-related outcomes.

4.5. Strengths and Limitations

This study has several strengths. It included a relatively large multi-region sample and applied a standardized pre-test and post-test framework, allowing assessment of immediate knowledge changes across diverse educational contexts. In addition, the combination of descriptive, inferential, adjusted, and sensitivity analyses provided a more comprehensive evaluation than a single pooled estimate. This approach is particularly relevant because environmental education outcomes may vary substantially across settings and populations [33].

Several limitations should be acknowledged. First, the study used a single-group pre-post design without an untreated comparison group; therefore, causal effects cannot be established. Alternative explanations, including testing effects, regression to the mean, maturation, response-shift bias, and concurrent external influences, cannot be completely excluded. Second, the evaluation focused on immediate knowledge change and did not assess knowledge retention, behavioral change, or longer-term sustainability-related outcomes.

Third, the assessment instrument was intentionally brief to facilitate standardized implementation across multiple school sites within a one-day educational activity. However, the limited number of items and the use of non-identical pre-test and post-test questions restricted evaluation of measurement reliability; therefore, the observed gains should be interpreted as changes in targeted IAQ-related knowledge rather than as comprehensive

measures of students' overall competence. Finally, regional grade distributions were imbalanced, and the analytic dataset lacked consistent teacher, class, or school identifiers for all participants, limiting the interpretation of subgroup effects and preventing a full multilevel implementation analysis. Sensitivity analyses, including exclusion of potentially problematic records and cluster-robust models, were therefore conducted to evaluate the robustness of the main findings. Importantly, these analyses should not be interpreted as confirming that the excluded records were invalid, as the dataset did not allow differentiation between true zero scores, nonresponse, coding errors, or administration-related problems at the individual-record level.

5. Conclusions

This multi-region program evaluation found that a standardized STEM-based IAQ sustainability education activity was associated with significant immediate knowledge gain overall, but also with substantial variation across contexts. The strongest gains were observed in Nepal and Indonesia, whereas Guangxi, China showed no significant positive improvement under the current intervention format. Baseline readiness was a central predictor of gain, but it did not fully explain all regional patterns.

For IAQ and sustainability education, the main implication is that curricula should not be assumed to transfer uniformly across settings. Effective program design requires baseline-informed differentiation, locally meaningful examples, and evaluation strategies that account for uneven starting points and implementation contexts. Future research should use comparison groups, delayed post-tests, behavioral and attitudinal outcomes, item-level diagnostics, and stronger school- or classroom-level implementation measures to determine whether short-term knowledge gains translate into durable learning and sustainability-relevant practice, including pro-environmental behavior change [34].

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/su18147165/s1>, Table S1: Example IAQ knowledge items used in the pre-test and post-test; Table S2: Summary of implementation sources; Table S3: Sensitivity analysis excluding records with post-test score = 0 and pre-test score > 0; Figure S1: Individual knowledge-gain distributions by anonymized implementation source; Figure S2: OLS coefficients under conventional and implementation-source clustered uncertainty; Figure S3: Leave-one-implementation-source-out robustness of the baseline-score coefficient; Figure S4: Distribution of records with post-test score = 0 and pre-test score > 0 by anonymized implementation source.

Author Contributions: Conceptualization, W.-J.D. and J.C.-K.L.; methodology, W.-J.D. and J.C.-K.L.; software, J.S.; validation, J.S., W.-J.D. and J.C.-K.L.; formal analysis, J.S.; investigation, W.-J.D. and J.S.; resources, W.-J.D.; data curation, J.S.; writing—original draft preparation, J.S.; writing—review and editing, W.-J.D., W.H. and J.C.-K.L.; visualization, J.S.; supervision, W.-J.D. and J.C.-K.L.; project administration, W.-J.D.; funding acquisition, W.-J.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Development ICS Research Fund (02224), the Education University of Hong Kong.

Institutional Review Board Statement: The study was conducted under university ethics approval through The Education University of Hong Kong (No. 2024/25-0233; 17 December 2024). Permission was also obtained from participating schools, and written consent was obtained from parents or guardians of minor participants prior to data collection.

Informed Consent Statement: Informed consent was obtained from the parents or legal guardians of all participating students, and student assent was obtained where applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Toyinbo, O. Indoor environmental quality, pupils' health, and academic performance—A literature review. *Buildings* **2023**, *13*, 2172. [\[CrossRef\]](#)
2. Branco, P.T.B.S.; Sousa, S.I.V.; Dudzińska, M.R.; Ruzgar, D.G.; Mutlu, M.; Panaras, G.; Papadopoulos, G.; Saffell, J.; Scutaru, A.M.; Struck, C.; et al. A review of relevant parameters for assessing indoor air quality in educational facilities. *Environ. Res.* **2024**, *261*, 119713. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Bozzola, E.; Agostiniani, R.; Pacifici Noja, L.; Park, J.; Lauriola, P.; Nicoletti, T.; Taruscio, D.; Taruscio, G.; Mantovani, A. The impact of indoor air pollution on children's health and well-being: The experts' consensus. *Ital. J. Pediatr.* **2024**, *50*, 69. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Kurmanbekova, M.; Du, J.; Sharples, S. A review of indoor air quality in social housing across Low-and Middle-Income countries. *Appl. Sci.* **2025**, *15*, 1858. [\[CrossRef\]](#)
5. Tran, V.V.; Park, D.; Lee, Y.C. Indoor air pollution, related human diseases, and recent trends in the control and improvement of indoor air quality. *Int. J. Environ. Res. Public Health* **2020**, *17*, 2927. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Roche, I.V.; Ubalde-Lopez, M.; Daher, C.; Nieuwenhuijsen, M.; Gascon, M. The health-related and learning performance effects of air pollution and other urban-related environmental factors on school-age children and adolescents—A scoping review of systematic reviews. *Curr. Environ. Health Rep.* **2024**, *11*, 300–316. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Boso, A.; Oltra, C.; Álvarez, B.; Garrido, J.; Hofflinger, Á.; Gálvez-García, G. Why do we misperceive air pollution? A scoping review of key judgmental biases. *Air Qual. Atmos. Health* **2025**, *18*, 447–460.
8. Adlong, W.; Dietsch, E. Environmental education and the health professions: Framing climate change as a health issue. *Environ. Educ. Res.* **2015**, *21*, 687–709.
9. Torkar, G.; Bogner, F.X. Environmental values and environmental concern. *Environ. Educ. Res.* **2019**, *25*, 1570–1581. [\[CrossRef\]](#)
10. Braun, T.; Cottrell, R.; Dierkes, P. Fostering changes in attitude, knowledge and behavior: Demographic variation in environmental education effects. *Environ. Educ. Res.* **2018**, *24*, 899–920.
11. Schmitz, G.L.; Rocha, J.B.T. Environmental education program as a tool to improve children's environmental attitudes and knowledge. *Education* **2018**, *8*, 15–20.
12. Maurer, M.; Koulouris, P.; Bogner, F.X. Green awareness in action—How energy conservation action forces on environmental knowledge, values and behaviour in adolescents' school life. *Sustainability* **2020**, *12*, 955. [\[CrossRef\]](#)
13. Demirci, S.; Reid, A.; Teksöz, G.; Sahin, E. Systems literacy: Towards a conspectus and model for environmental and sustainability education. *Environ. Educ. Res.* **2025**, *31*, 856–891. [\[CrossRef\]](#)
14. Liefländer, A.K.; Bogner, F.X. Educational impact on the relationship of environmental knowledge and attitudes. *Environ. Educ. Res.* **2018**, *24*, 611–624.
15. Deng, W.-J.; Sun, J.; Ho, W.; Lee, J.C.-K. Cross-national evaluation of STEM-based indoor air quality (IAQ) education for sustainable development. *Aerosol Air Qual. Res.* **2026**, *in press*. [\[CrossRef\]](#)
16. Li, Q.; Jiang, J.; Wang, S.; Rumchev, K.; Mead-Hunter, R.; Morawska, L.; Hao, J. Impacts of household coal and biomass combustion on indoor and ambient air quality in China: Current status and implication. *Sci. Total Environ.* **2017**, *576*, 347–361. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Malla, S. Household transitions to clean energy from traditional biomass in Nepal: Challenges and opportunities. *Nepal Public Policy Rev.* **2021**, *1*, 48–67. [\[CrossRef\]](#)
18. Zhou, R.; Lee, N. The reception of education for sustainable development (ESD) in China: A historical review. *Sustainability* **2022**, *14*, 4333. [\[CrossRef\]](#)
19. Ghimire, S.; Mishra, S.R.; Sharma, A.; Siweya, A.; Shrestha, N.; Adhikari, B. Geographic and socio-economic variation in markers of indoor air pollution in Nepal: Evidence from nationally-representative data. *BMC Public Health* **2019**, *19*, 195. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Vakalopoulos, A.; Dharmage, S.C.; Dharmaratne, S.; Jayasinghe, P.; Lall, O.; Ambrose, I.; Weerasooriya, R.; Bui, D.S.; Yasaratne, D.; Heyworth, J.; et al. Household air pollution from biomass fuel for cooking and adverse fetal growth outcomes in rural Sri Lanka. *Int. J. Environ. Res. Public Health* **2021**, *18*, 1878. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Norazman, N.; Che Ani, A.I.; Ismail, W.N.W.; Hussain, A.H.; Abdul Maulud, K.N. Indoor environmental quality towards classrooms' comforts level: Case study at Malaysian secondary school building. *Appl. Sci.* **2021**, *11*, 5866. [\[CrossRef\]](#)
22. Karim, N.; Othman, H.; Zaini, Z.I.I.; Rosli, Y.; Wahab, M.I.A.; Al Kanta, A.M.; Omar, S.; Sahani, M. Climate change and environmental education: Stance from science teachers. *Sustainability* **2022**, *14*, 16618. [\[CrossRef\]](#)
23. Gao, J.; Jiang, Y. Development and validation of China's Environmental Citizenship Questionnaire: Evidence from junior high school students. *Front. Environ. Sci.* **2024**, *12*, 1105469. [\[CrossRef\]](#)

24. Liang, Z.; Zou, Y.; Xu, C.; Chen, J. The Effects of Environmental Education on Residents' Ecological Security Behavior: The Mediating Role of Nature's Psychological Ownership Perspective and The Moderating Role of Visual Fluency. *Environ. Manag.* **2024**, *73*, 338–353.
25. Rawat, N.; Kumar, P. Interventions for improving indoor and outdoor air quality in and around schools. *Sci. Total Environ.* **2023**, *858*, 159813. [[CrossRef](#)] [[PubMed](#)]
26. Stapleton, S.R.; Reif, K. Teaching outside as third space: Toward school science that acknowledges student ecological expertise. *Environ. Educ. Res.* **2022**, *28*, 1373–1390. [[CrossRef](#)]
27. Van De Wetering, J.; Leijten, P.; Spitzer, J.; Thomaes, S. Does environmental education benefit environmental outcomes in children and adolescents? A meta-analysis. *J. Environ. Psychol.* **2022**, *81*, 101782. [[CrossRef](#)]
28. Aeschbach, V.M.J.; Schwichow, M.; Rieß, W. Effectiveness of climate change education—A meta-analysis. *Front. Educ.* **2025**, *10*, 1563816. [[CrossRef](#)]
29. Karanth, K.K.; Unnikrishnan, S.; Salazar, G. From classrooms to conservation: Scaling environmental education across India's Western Ghats. *Front. Conserv. Sci.* **2025**, *6*, 1659491. [[CrossRef](#)]
30. Ongwandee, M.; Khianthongkul, K.; Panyametheekul, S.; Yongprapat, K.; Srinaka, K.; Morris, J. Bangkok school indoor air quality: Monitoring and intervention by positive pressure fresh air system. *Environ. Sci. Pollut. Res.* **2024**, *31*, 25454–25467. [[CrossRef](#)]
31. Therivel, R.; González, A. "Ripe for decision": Tiering in environmental assessment. *Environ. Impact Assess. Rev.* **2021**, *87*, 106520. [[CrossRef](#)]
32. Zhai, J.; Dillon, J.; Dawson, V. Climate Change Education: Cross-Cultural Perspectives. *ECNU Rev. Educ.* **2025**, *8*, 49–52.
33. Ballard, H.L.; Lindell, A.J.; Jadallah, C.C. Environmental education outcomes of community and citizen science: A systematic review of empirical research. *Environ. Educ. Res.* **2024**, *30*, 1007–1040. [[CrossRef](#)]
34. Świątkowski, W.; Surret, F.L.; Henry, J.; Buchs, C.; Visintin, E.P.; Butera, F. Interventions promoting pro-environmental behaviors in children: A meta-analysis and a research agenda. *J. Environ. Psychol.* **2024**, *96*, 102295. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.