Teaching atmospheric reasoning with interactive model anatomy: a framework and evaluation protocol

Adrian Dunkley
Climate Studies Group Mona
The University of the West Indies, Mona
adrian.dunkley@mymona.uwi.edu

Draft v0.1 — 31 October 2026
Theory and evaluation protocol; not peer reviewed or submitted

Abstract

This paper specifies an instructional framework for teaching conservation, observation support and uncertainty through manipulable isometric instruments linked to climate-physics papers. The framework requires learners to predict a consequence, test a controlled change, explain the mechanism and apply it in an unfamiliar case. An engaging animation is treated as an instructional resource, not evidence of learning. The primary proposed outcome is delayed, unaided transfer to new scientific reasoning tasks. Confidence, enjoyment, immediate assisted success and climate-action participation are measured separately. A pilot checks language, accessibility, scientific misconceptions and assessment reliability before any comparative study. The subsequent design compares matched static and interactive materials, with a separate scaffolded-interaction arm if recruitment and teaching conditions permit. It extends the author's ClimateQuest and preparedness work while keeping sport participation, scientific understanding and verified community action as distinct constructs. The contribution is the learning framework, assessment specification and study protocol; no participant study or learning-effect estimate has been completed for this paper. Its thesis relationship is educational translation: the same distinctions used to evaluate atmospheric reconstruction become assessable learning goals, without presenting a simplified visual instrument as a complete atmospheric solver.

Plain-language goal. Help learners predict, test and explain a climate mechanism, then check whether they can use it later without assistance.

Thesis relationship. Turns conservation, observation support and uncertainty into assessable reasoning tasks. Educational translation is separate from the physical-model contribution.

1. The scientific reasoning to be taught

The thesis depends on distinctions that are easy to lose in a polished model visualisation. A shared transfer can conserve a total while moving the wrong amount. A finer grid can look detailed without recovering independent information. A reconstruction can use observations that a forecast would not yet have had. An uncertainty interval can be narrow because the model omits relevant variation. The education framework turns those distinctions into tasks a learner must answer and explain.

The central research question is whether controlled interaction improves delayed transfer beyond matched static explanation. The proposed audience begins with learners able to work with ratios, graphs and simple algebra. A comprehension pilot determines the actual age range and reading level; the label ELI15 is an accessibility goal, not evidence that every fifteen-year-old will understand the material. Local teachers and learners help identify unfamiliar terms and inappropriate examples.

Climate education research motivates personally meaningful material and active engagement, while leaving the effectiveness of this particular resource untested [1]. Broader STEM evidence supports studying active learning, but it cannot supply an effect size for this setting [2]. The study therefore measures its own outcomes rather than importing a benefit from another population.

2. Instructional framework and mechanism

Each lesson follows four moves. First, the learner commits to a prediction before the instrument moves. Second, one scientifically meaningful control changes. Third, the learner explains why the outcome occurred and identifies a limit of the model. Fourth, a changed case tests whether the explanation transfers beyond the original display. The sequence prevents success from depending entirely on watching the answer appear.

The isometric object must embody the relationship being taught. Two reservoir chambers show a shared transfer because their total can be checked directly. A sampling microscope shows the distinction between output mesh and available observations. A mask-and-scan instrument distinguishes visible context from a hidden target. These objects are original teaching analogies built with the Anatomy drawing kit; their appeal is not itself a causal educational mechanism.

Simulation research supports attention to instructional support and reflection, rather than assuming that interaction always improves learning [3]. Multimedia-learning theory motivates reducing irrelevant visual and verbal load [4]. The design consequently keeps one scientific question on each instrument, puts explanations outside the drawing and removes repeated captions that merely describe visible motion.

3. A conservation lesson with an explicit limit

Let two constructed reservoirs contain m_A and m_B units. A chosen transfer q gives:

$$m'_A = m_A - q, \qquad m'_B = m_B + q, \qquad m'_A + m'_B = m_A + m_B.$$

The learner predicts both final amounts and their total. A second case uses an inconsistent pair of transfers, so the total changes. A third case adds a stated external source, requiring a revised budget. The key assessment is an explanation of which changes are internal and which cross the system boundary.

The lesson also asks whether a conserved total proves that the transfer was physically correct. It does not. A wrong q can preserve the same sum. Likewise, an excessive transfer can make one amount negative unless the update rule restricts it. These counterexamples matter because the thesis separates conservation, positivity, physical identification and predictive usefulness. The simplified instrument makes that reasoning inspectable without claiming to solve atmospheric transport.

A transfer task then changes the surface story: heat exchange, water sharing or material inventory

replaces the original chambers. The numbers and diagram layout change as well. A learner who memorised one slider position should not pass through recognition alone. Open explanations are scored with a rubric that distinguishes a correct number from a correct account of the mechanism.

4. Observation support and uncertainty lessons

The observation lesson starts with one area-averaged value and several possible fine-scale arrangements that share it. The learner identifies why the average cannot select a unique arrangement. Additional independent observations then constrain the possibilities. A detailed rendered field is labelled as one candidate, preventing the display from teaching that higher pixel density automatically means higher physical resolution.

The uncertainty lesson compares a tightly fitted wrong model with a wider distribution that represents additional uncertainty. It asks what information would justify narrowing the interval. It does not teach that wider is always better. Scoring rewards the ability to connect a statement of uncertainty to its source, rather than reciting that all models are uncertain. Scientific content is aligned to established physical-science explanations, with the specific local curriculum alignment still to be reviewed [5, 6].

The paper's educational contribution is conceptual reasoning, not a general claim about computer simulations. A circuit-simulation study offers a relevant example of measuring transfer to a real task, while its domain and learners differ from this study [7]. The proposed assessments therefore include unfamiliar contexts and explanations, rather than relying solely on performance inside the animation.

5. Relationship to ClimateQuest and preparedness work

ClimateQuest already proposes sport-linked lessons, verified community activity and locally directed fundraising. Before the Storm proposes a separate preparedness pathway. This framework supplies a possible scientific-reasoning component for those resources. It does not relabel participation, donated money or a completed activity as evidence of scientific competence.

The initial study can be conducted as a standalone lesson before integration with a club or campaign. That reduces ambiguity about whether effects arise from the instrument, competition, teacher attention, sport participation or group identity. A later implementation study may examine those additional mechanisms with explicit outcomes and comparison groups. The existence of an attractive poster or functioning website does not establish that a pilot occurred.

Community action requires an independent verification rule. An uploaded photograph may establish that a task was attempted without establishing environmental benefit. Activities involving physical hazards require supervision and locally appropriate permissions. No participant records or community claims have been collected for the present protocol.

6. Accessible delivery and the role of AI

Learners must be able to pause the animation, use a keyboard, choose reduced motion and read a static explanation. Printed diagrams and low-bandwidth alternatives provide the same scientific

relationship. Universal Design for Learning supplies a design reference for multiple representation and response modes [8]. Accessibility is checked with actual users; the presence of an alt attribute alone does not establish usability.

If an AI assistant is included, its role is limited to clarifying approved explanations, asking the learner to justify a step and identifying uncertainty. The assessed delayed task is unaided. Store whether assistance was available and used, since assisted success and retained competence answer different questions. UNESCO guidance motivates human oversight and review of data and age-related conditions [9]. It does not establish permission to collect children's conversations.

The curriculum also separates learning about climate mechanisms from prescribing one behaviour. UNESCO climate-action curriculum guidance is a candidate alignment resource [10]. Teachers and communities must review its fit with their context. Learners can demonstrate scientific reasoning while disagreeing about a policy choice whose values or costs remain contested.

7. Pilot and comparative study

The pilot uses cognitive interviews and observed lesson completion to test wording, control interpretation, visual ambiguity and the assessment rubric. Ask learners to explain the reservoir boundary, the source of a gridded value and what a model could have known at issue time. Record misconceptions introduced by the instrument itself. Revise the resource before freezing the comparative version; do not count those revisions as evidence of efficacy.

The proposed comparison has a static-diagram condition and an interactive condition with matched content, teaching time and examples. A third condition adds prediction-and-explanation scaffolding if the study can distinguish its contribution without underpowering the main comparison. The simpler two-condition design is preferable when recruitment or teaching capacity is limited. Sample size follows pilot variability and a declared minimum meaningful difference, rather than an invented enrolment target.

Randomisation occurs at learner or class level according to contamination risk and teaching arrangements. Class assignment requires cluster-aware design and reporting [11]. The analysis plan is checked against educational-study standards before recruitment [12]. Teachers cannot silently compensate one condition with additional examples, and outcome assessors should not know which resource a learner received when scoring explanations.

The primary assessment is delayed and unaided. An immediate post-test is secondary, and a baseline measures prior reasoning. Retrieval-practice research motivates the delayed assessment while cautioning that repeated assessment can itself influence learning [13, 14]. Use parallel forms and balance exposure across conditions. Record attrition, absences and reasons for missing outcomes without treating missing learners as zero scorers.

8. Outcomes and analysis

The primary outcome is a prespecified transfer score built from novel problems and explanation items. The rubric distinguishes recognising the relevant system boundary, applying the mathematical relationship and identifying a material limit. Items undergo expert scientific review and learner comprehension testing. A total score is not described as a validated measure until reliability and construct evidence support that claim.

A baseline-adjusted comparison may be expressed as:

$$Y_{ic}^{\text{delayed}} = \alpha + \tau Z_c + \beta Y_{ic}^{\text{baseline}} + u_c + \epsilon_{ic}.$$

Here Z_c indicates the assigned condition when randomisation occurs by class, u_c represents class-level variation and τ is the intended assignment effect under the stated design. The model, variance estimator, missing-data approach and any small-cluster limitation must be fixed before outcome inspection. Learner-level randomisation requires the corresponding assignment structure. Report uncertainty and item-level patterns alongside the aggregate.

Confidence, enjoyment, self-reported understanding, time on task, assistance use and verified action remain secondary and separate. Research comparing felt and actual learning motivates this separation [15]. A learner can enjoy the resource and misunderstand conservation, or find it demanding while learning more. Neither confidence nor satisfaction is used as a substitute for an independently scored task.

9. Ethics, failure criteria and interpretation

The study requires appropriate institutional review, teacher agreement and participant consent or guardian permission where applicable. These approvals have not been obtained by writing the protocol. Recruitment must avoid making participation a condition for grades, club membership or access to learning resources. Data minimisation, retention, withdrawal and reporting rules are agreed before collection.

The resource fails its scientific-design gate if it suggests that conserved totals prove accurate physics, that latent predictions are observations or that detailed animation establishes forecast skill. The evaluation fails its inference gate if conditions differ materially in teaching time, if assistance enters the unaided outcome or if the analysis ignores the assignment structure. A null or adverse comparison remains informative and must be retained.

Transfer beyond the sampled learners, language and teaching conditions requires further evidence. The paper does not presume that younger learners benefit more from interactivity or that one instructional style fits all Caribbean classrooms. The relevant question is which mechanism, support and delivery condition helps a specified reasoning task.

10. Current contribution and available prework

Available prework comprises the author's ClimateQuest abstract and posters, preparedness materials, the climate-physics paper series and existing interactive mathematical explanations. The current contribution specifies how those materials become a testable educational intervention. The web instruments are usable learning resources, but there are no participant results to report.

The next checkpoint is a scientifically reviewed lesson set, parallel assessment forms, a comprehension pilot and an agreed ethics route. The framework then supports a controlled study of delayed unaided reasoning. Its relationship to the physics thesis is the faithful translation of evidence, uncertainty and model limits into concepts that learners can test and explain.

Data, authorship and assistance

Adrian Dunkley is the sole named author. The climate-research context is Climate Studies Group Mona, The University of the West Indies, Mona; ClimateQuest's educational context is the Caribbean AI Association. These affiliations imply no institutional endorsement or ethics approval. Contact: adrian.dunkley@mymona.uwi.edu. AI assisted literature organisation, drafting and the interactive instruments. The author must verify references, scientific examples, funding, conflicts and declarations before external submission. Source-review depth is recorded separately. No learner data, achieved learning effect, completed trial or educational certification is asserted.

References

[1] Monroe, Plate, Oxarart, Bowers and Chaves (2019). Identifying effective climate change education strategies: a systematic review of the research. https://doi.org/10.1080/13504622.2017.1360842.

[2] Freeman et al. (2014). Active learning increases student performance in science, engineering, and mathematics. https://doi.org/10.1073/pnas.1319030111.

[3] Smetana and Bell (2012). Computer Simulations to Support Science Instruction and Learning: A critical review of the literature. https://doi.org/10.1080/09500693.2011.605182.

[4] Mayer and Moreno (2003). Nine Ways to Reduce Cognitive Load in Multimedia Learning. https://doi.org/10.1207/S15326985EP3801_6.

[5] IPCC (2021). Climate Change 2021: The Physical Science Basis. Summary for Policymakers. https://www.ipcc.ch/report/ar6/wg1/chapter/summary-for-policymakers/.

[6] National Research Council (2012). A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas. https://doi.org/10.17226/13165.

[7] Finkelstein et al. (2005). When learning about the real world is better done virtually: A study of substituting computer simulations for laboratory equipment. https://doi.org/10.1103/PhysRevSTPER.1.010103.

[8] CAST (2024). Universal Design for Learning Guidelines, version 3.0. https://udlguidelines.cast.org/.

[9] Miao and Holmes / UNESCO (2023). Guidance for generative AI in education and research. https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research.

[10] UNESCO (2024). Greening curriculum guidance: Teaching and learning for climate action. https://www.unesco.org/en/articles/greening-curriculum-guidance-teaching-and-learning-climate-action.

[11] Campbell, Piaggio, Elbourne and Altman (2012). Consort 2010 statement: extension to cluster randomised trials. https://doi.org/10.1136/bmj.e5661.

[12] Institute of Education Sciences (2022). What Works Clearinghouse Procedures and Standards Handbook, Version 5.0. https://ies.ed.gov/ncee/wwc/Handbooks.

[13] Roediger and Karpicke (2006). Test-enhanced learning: taking memory tests improves long-term retention. https://doi.org/10.1111/j.1467-9280.2006.01693.x.

[14] Dunlosky et al. (2013). Improving Students' Learning With Effective Learning Techniques: Promising Directions From Cognitive and Educational Psychology. https://doi.org/10.1177/1529100612453266.

[15] Deslauriers et al. (2019). Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. https://doi.org/10.1073/pnas.1821936116.

Reference review record

The following record distinguishes primary abstract checks from proposed references requiring additional reading. It is not a statement that all full texts were verified.

monroe Publisher abstract inspected; relevance, engagement and discussion motivation.

freeman Primary PubMed abstract inspected; undergraduate STEM synthesis, not adolescent Caribbean validation.

smetana Publisher abstract inspected through search; instructional scaffolding matters.

mayer Publisher abstract inspected through search; cognitive-load design motivation.

wgi Proposed scientific-content reference; does not establish learning effectiveness.

k12 Publisher retrieval rate limited; proposed curriculum-alignment reference.

finkelstein Publisher abstract inspected; circuits, not climate education.

udl Official guidelines inspected; accessible representation and response modes.

genaied Official UNESCO publication page inspected; 2023 guidance; data, age and teacher oversight require local review.

greening Official UNESCO publication page inspected; 2024 guidance, page updated June 2026; local alignment untested.

cluster Publisher retrieval blocked; reporting reference, full-text check pending.

wwc Official handbook page inspected; version 5.0, August 2022 revised December 2022; actual study design still requires review.

testing Primary PubMed abstract inspected; delayed retention motivation.

dunlosky Primary PubMed abstract inspected; practice and evaluation rationale.

deslauriers Primary PubMed abstract inspected; self-report and measured learning separated.