VOLUME 01 / 2026

Designership Press

Learning That Can Answer Back

Dr. Freedom Cheteni

Original scholarship in K-12 learning, human agency and the design of consequential work.

Research edition 1.0 | Prepared 2026-10-16

Contents

Learning That Can Answer Back

From Virtual Presence to Physical Competence: Consequence-Preserving Transfer in K-12 Spatial STEM 4

Who Holds the Classroom? Bounded Pedagogical Autonomy for Social Robots and AI Agents in K-12 Education 17

Public Contribution Without Prestige Proxies: Designership as a Mentored Inquiry Pathway in K-12 Education 32

A School Must Be Reachable: Digital Infrastructure, Accessible Opportunity, and the Designership Service Obligation 46

From Paper Agents to Accountable Student Scholarship: A Claim-Evidence-Contribution Ledger for K-12 Research 60

Each article is available as a complete reading edition with its own source-grounded paper agent and downloadable PDF. Page numbers refer to this collected volume.

Editorial and research record

Research record

The Research University | Volume 1, 2026 | Edition 1.0

Original analyses and prospective protocols. New independent human peer review is not recorded for this edition.

No DOI is assigned to this research edition. Existing identifiers in references identify the cited source works.

https://journal.research.university/press/designership/volumes/1

Author ORCID: 0009-0006-3507-0094

Editorial record: https://journal.research.university/press/designership/editorial

This volume develops original, evidence-informed arguments and prospective studies from the Designership research agenda. Conceptual contributions are distinguished from completed empirical findings. The study protocols describe work to be registered, approved and conducted; they are not reports of completed trials.

Human authority and transparent assistance

Dr. Freedom Cheteni is the human author responsible for the scholarship and its final approval. AI assistance supported research discovery, manuscript development and editorial checks. AI tools are neither authors nor independent human peer reviewers. The linked editorial record identifies what was checked and what remains subject to independent human judgement.

Citation and reuse

Cited research retains its original attribution, venue and identifiers. These new articles do not replace the original publications or transfer their authorship. Readers should use the version and source record associated with each claim. This volume assigns no new DOI and makes no external institutional endorsement claim.

D01 / Spatial competence

From Virtual Presence to Physical Competence: Consequence-Preserving Transfer in K-12 Spatial STEM

Dr. Freedom Cheteni

The Research University | Volume 1, 2026 | Edition 1.0 | D01

Critical integrative analysis and prospective research protocol

Original analyses and prospective protocols. New independent human peer review is not recorded for this edition.

DOI: not assigned to this research edition.

Read online | Author ORCID: 0009-0006-3507-0094 | Editorial record

Abstract

Spatial computing can make science vivid without establishing that learning remains useful outside a simulation. This critical integrative analysis develops consequence-preserving transfer as a testable extension of Freedom Cheteni's proposed Movement Thinking framework. The contribution combines a consequence map, an evidence-linked revision record, and an explicit transfer contract. It distinguishes meaningful control over scientific consequences from display immersion, motor activity, and reported presence. A prospective classroom-cluster randomised trial would compare a structured prediction-action-revision sequence with an equally resourced active curriculum, offering accessible immersive and non-headset routes in both arms. The primary estimand is the intention-to-treat effect on independently scored, delayed physical transfer among baseline-enrolled eligible participants. The protocol specifies allocation, outcome validation, power simulation, clustering, missingness, participant protections, and falsification criteria. No primary data or efficacy results are reported. The resulting research agenda treats accessibility, evidential limits, and the capacity to act competently after a device is removed as central conditions of educational value.

Keywords: Movement Thinking, Designership, spatial STEM, extended reality, transfer of learning, accessible design, cluster randomised trial, K-12

1. The educational claim begins outside the headset

A student who completes a virtual laboratory has demonstrated competence in that environment. Whether the student can reason, act safely, and revise a design in a physical laboratory remains a separate question. That distinction matters for K-12 systems investing in spatial computing. Attractive environments, confident learners, and polished demonstrations can be educationally valuable, but none establishes transfer. A signature Designership research programme should therefore begin with a demanding criterion: students must use what they learned when the display, objects, interaction controls, and social situation change.

This article develops consequence-preserving transfer as a proposed extension of Freedom Cheteni’s Movement Thinking framework. The contribution is an inspectable design logic: preserve the consequential relationship between a learner’s decision, an observable outcome, and a justified revision while allowing the interface and physical demands to vary. An experiment follows from that logic. The relevant intervention is not simply access to a headset; it is a curriculum that requires prediction, action, comparison, explanation, and subsequent application to unfamiliar physical materials.

The argument concerns secondary STEM education, particularly introductory engineering and mechanics. It does not claim that existing VR School practice has already produced the effects proposed here. No participants were recruited, primary outcomes collected, or trial results estimated for this article. The manuscript combines a critical integrative analysis with a prospective research protocol. Its value lies in specifying claims that could fail, documenting how they would be evaluated, and making an accessible alternative part of the intervention rather than an exception to it.

2. Evidence selection and limits of this analysis

The analysis uses a deliberately bounded set of foundational empirical studies, a theory of immersive learning, a transfer taxonomy, and authoritative accessibility and research-governance documents. Searches conducted on 16 October 2026 used combinations of immersive virtual reality, embodied learning, physical transfer, science learning, high school, mixed reality, accessibility, and cluster randomised trial. Publisher pages and author-institution repositories were preferred; DOI metadata were checked against Crossref. The selection was purposive and question-led. It was not an exhaustive database search, screened systematic review, quantitative synthesis, or estimate of average XR effectiveness.

Sources were selected to test a tension rather than accumulate endorsements: action can contribute to learning, while immersion can add demands without improving knowledge. The analysis distinguishes college physics, whole-body mixed reality, and high-school virtual science studies rather than treating their populations and technologies as interchangeable. Published experimental effects support the

plausibility of mechanisms within their original settings. They do not validate Cheteni's framework, determine a school district's expected benefit, or supply a transferable effect size for the proposed trial.

3. What embodiment supports-and what it leaves unproven

Kontra, Lyons, Fischer, and Beilock investigated physical experience in science learning through laboratory experiments and a college physics field experiment. Their work supports taking action-relevant experience seriously when teaching a physical concept; it does not establish that any movement, headset, or immersive scene improves K-12 learning. A student can move vigorously while attending to the wrong relationship. The instructional question is whether action makes a consequential scientific distinction available for inspection. [R1]

Lindgren, Tscholl, Wang, and Johnson examined embodied interaction within a mixed-reality simulation. Their study provides a closer precedent for designing movement around a scientific relation rather than adding movement as decoration. Johnson-Glenberg's educational design account similarly makes gesture and manipulation central considerations. These contributions motivate a design commitment, not a universal prescription: the permitted action should correspond to the concept the learner needs to understand. [R2, R3]

Consider balancing an unfamiliar beam. Rotating a virtual model with a controller is an interface action. Predicting how moving a load changes balance, observing an unexpected result, and revising the prediction is epistemic work. A seated learner using a switch or keyboard may undertake that same epistemic work. Consequently, this article locates the proposed mechanism in accessible control over meaningful consequences, not in a required body movement. Physical realism is useful only where it preserves information needed for the target task.

4. Presence is a process variable, not the learning verdict

Makransky, Terkildsen, and Mayer compared head-mounted and desktop science simulations with university students. Greater reported presence accompanied poorer learning in that experiment. The result is not a verdict against all immersive learning; it is a direct counterexample to treating presence as sufficient evidence of educational value. A district can produce an impressive experience while making the central explanatory work harder. [R4]

More directly relevant to secondary schooling, Makransky, Andreasen, Baceviciute, and Mayer studied high-school students using immersive science simulations or video. Immersion did not by itself provide the reported learning advantage, while post-lesson enactment benefited procedural knowledge and transfer in the immersive condition. This finding motivates testing what learners do around the simulation, including reconstruction and explanation, rather than evaluating a headset as an isolated

treatment. It does not establish the effectiveness of the beam-and-balance curriculum proposed below. [R5]

The Cognitive Affective Model of Immersive Learning provides a framework for separating technological features, presence and agency, cognitive and affective processes, and different learning outcomes. This manuscript adopts that separation while narrowing its causal claim to a particular instructional sequence and physical transfer task. Enjoyment and presence will be measured as secondary process variables. They will never substitute for competent performance, justify excluding uncomfortable learners, or determine the decision to scale. [R6]

5. Consequence-preserving transfer as a Movement Thinking proposition

Movement Thinking is used here as Cheteni's proposed orientation toward visible questions, bounded authority, preserved revision, and accountability for consequences. Consequence-preserving transfer makes that orientation experimentally specific. It predicts that learners will transfer more successfully when they repeatedly commit to a prediction, control a relevant variable, inspect the outcome, explain a discrepancy, and test a revision across representations. Each link is necessary for the instructional claim; a simulation that supplies the explanation before a student makes a prediction is a different intervention.

The proposed contribution has three components. First, a consequence map states which physical relationships a representation preserves and which it omits. Second, a revision record connects the learner's initial prediction, observed outcome, and changed explanation. Third, a transfer contract defines the unfamiliar task on which learning must remain useful. Together these components make the lesson's inferential promise visible. The map concerns model fidelity, the record concerns reasoning over time, and the contract concerns the boundary of generalisation.

For a balance task, the consequence map would preserve load magnitude, distance from the pivot, and direction of rotation while explicitly omitting material deformation and friction unless they are taught. The revision record would capture a prediction before simulation feedback and a reason after it. The transfer contract would require balancing a differently configured physical apparatus and diagnosing a misleading visual arrangement. This specification prevents a strong score on an identical virtual puzzle from being presented as broad engineering competence.

Barnett and Ceci's transfer taxonomy helps identify why changing environments is not a single distance. Content, context, timing, function, and modality can vary together. The proposed curriculum therefore changes representation while controlling the scientific relationship, then evaluates limited transfer across changed materials and conditions. It does not claim far transfer to general intelligence, employability,

creativity, or all future STEM learning. Such outcomes would require their own operational definitions and studies. [R7]

6. Accessibility is part of the treatment definition

The W3C XR Accessibility User Requirements identifies needs spanning navigation, interaction, presentation, and other aspects of immersive environments. It is a Working Group Note, not a certification that a particular lesson is accessible. Its relevance here is architectural: alternative inputs and outputs must be planned before recruitment and built into the curriculum specification. Compliance claims would require evaluation of the actual product and learners' experiences. [R8]

Every lesson would support seated participation, adjustable reach and movement amplitude, remappable inputs, captions, text equivalents, and a stable non-headset route with the same decision sequence. Instructions would avoid colour-only distinctions and timed controller movements as prerequisites for demonstrating understanding. Learners could request breaks or change routes without needing to disclose a diagnostic label to classmates. User testing would include students with relevant sensory, mobility, communication, and attentional access needs, compensated through an approved process.

Equivalent intellectual demand does not imply identical sensory experience. The study would audit whether each route exposes the learner to the same scientific relations and permits equally informative predictions and revisions. An independent accessibility review would identify unresolved barriers before trial commencement. Participation in research would never require a student to abandon an existing accommodation. Where access changes the scoring procedure, the change would be documented and assessed for construct equivalence rather than hidden inside an adjustment variable.

7. Prospective study: question, population, and allocation

The primary question is whether a consequence-preserving instructional sequence improves delayed physical transfer relative to an equally resourced simulation curriculum without the structured prediction-and-revision sequence. The proposed study is a parallel, two-arm, classroom-cluster randomised trial in grades 8-10. Both arms receive the same scientific content, total instructional time, teacher contact, equipment access, orientation, and physical assessment. The contrast concerns instructional organisation, not entitlement to resources.

Participating schools would be recruited across differing resource settings before teachers learn allocation. Eligible students would be enrolled in participating STEM classes and able to undertake the scientific decisions through at least one accessible route. Device discomfort or disability would not be an automatic exclusion. Clinical or equipment-specific safety concerns would be managed through a reviewed individual

access plan and a non-headset route. Research consent would be obtained separately from ordinary curriculum access. The final population definition and exclusion rules would be registered before assignment.

An independent statistician would generate allocation after class rosters, baseline measures, and teacher commitments are fixed. Classrooms would be randomised within schools and grade bands, balancing teacher blocks where feasible. If teachers teach both arms, scripted materials and observation would monitor cross-arm contamination; if spillover cannot be limited, schools rather than classes would become clusters, requiring a revised power plan before recruitment. Allocation would remain concealed from recruiters until baseline enrolment closes. Students and teachers cannot be blinded to the sequence; physical-task assessors and primary analysts would be masked to coded assignment.

The study would follow an adapted cluster-trial reporting plan: identify recruitment relative to allocation, explain cluster and student eligibility, describe the randomisation unit, report attrition at both levels, and account for clustering in analysis. These requirements draw on the CONSORT extension for cluster trials; the educational protocol would also identify departures arising from school scheduling and consent. No registration number, ethics approval, participating district, or committed recruitment site is claimed here. [R9]

8. Intervention specification and active comparator

The intervention would comprise four curriculum sessions over two weeks, followed by an immediate assessment and a delayed assessment four weeks after the final lesson. These are proposed design choices, not completed procedures. Sessions would begin with accessible orientation common to both arms. Learners would investigate balance and moments through parameter-controlled simulations and an identical allocation of hands-on materials. The intervention arm would record predictions before changing parameters, receive outcome feedback, explain discrepancies, and test one documented revision.

The active comparator would receive the same demonstrations, examples, simulation capabilities, hands-on time, and teacher explanations. Its practice sequence would use conventional worked examples and answer checks without the required prediction-outcome-revision record. It would include an equal-duration note activity so that writing time is not the principal contrast. The study would therefore estimate the added effect of the structured sequence over a credible alternative. It would not estimate an unqualified VR-versus-no-VR effect, nor isolate every component of Movement Thinking.

Both arms would offer a choice between compatible immersive and non-headset routes under the same accessibility policy. Route choice, switching, and exposure would be documented before and during lessons. Because route is not randomised,

differences between route users would be descriptive and potentially confounded. A later factorial trial could isolate immersion, but the present study prioritises the effect of an inclusive instructional policy. Teachers would use common stopping and support rules, and receive preparation in delivering both arms without enthusiasm or feedback asymmetries.

Fidelity would be assessed using a prespecified sample of lesson observations and de-identified work records. Observers would score whether the required decision preceded feedback, whether a consequence was visible, and whether an explanation supported revision. They would also document hints, departures from scripts, technical failures, and extra instruction. Fidelity would inform interpretation and future implementation; excluding low-fidelity classrooms from the primary estimate would undermine the policy question the trial is designed to answer.

9. Hypotheses, estimand, and outcomes

H1 predicts a positive average effect of assignment to the consequence-preserving sequence on delayed physical transfer. H2 predicts that the sequence improves explanation-supported revision rather than merely increasing the number of edits. H3 predicts that any primary benefit remains interpretable alongside access and tolerability: an advantage accompanied by systematic loss of participation or increased harm in an access-needs group cannot support an unrestricted adoption claim. H2 is secondary; H3 is a decision constraint assessed with uncertainty, not a promise that a small trial can prove subgroup equivalence.

The primary estimand is the student-average intention-to-treat effect of assignment among consenting, baseline-enrolled eligible students in the recruited clusters. It compares two instructional policies with their common access options. The primary endpoint is the total score on an independently administered physical transfer assessment four weeks after instruction. The scoring blueprint would be frozen before allocation: prediction under an unfamiliar load configuration, justified selection of a corrective action, diagnosis of a misleading configuration, and explanation of evidence supporting revision. Curriculum developers would not score their own students’ primary tasks.

Rubric dimensions would separate scientific correctness, evidential justification, and safe procedure. Task variants would alter materials and surface features while preserving the targeted scientific relationships. Expert review and a separate pilot would examine content coverage, reading burden, accessibility, and scoring reliability; pilot participants would not enter the confirmatory trial. Two masked assessors would independently score a prespecified subset, reconcile training disagreements before final scoring, and report agreement with uncertainty. The total score and weighting would not be redesigned after inspecting treatment differences.

Secondary outcomes would include immediate transfer, concept knowledge, revision quality, completion, adverse symptoms, route switching, and teacher workload. Revision quality would require a coded connection between evidence and a changed scientific explanation; edit counts alone would not qualify. Presence and enjoyment would be brief process measures with instrument versions and permissions established before use. Public artifacts would contain neither identifiable student videos nor eye, face, voice, or fine-grained movement traces. No biometric measure is necessary to test the principal educational claim.

10. Analysis, power, and missing observations

The primary analysis would estimate assignment effects using a model that includes the registered baseline task score and randomisation blocks and accounts for students clustered within classrooms. Where school-level randomisation is required, the model and uncertainty calculation would follow that level instead. The protocol would specify whether inference uses a multilevel model or cluster-robust estimation with an appropriate small-sample correction, chosen before outcomes are revealed. An allocation-consistent randomisation analysis would provide a sensitivity check where the number and arrangement of clusters permit it. Analysis will target equal weight for each baseline-enrolled eligible student, using prespecified individual weights, cluster-size handling, and marginal standardisation. Sensitivity analyses will assess informative cluster size rather than assume that a generic multilevel-model coefficient automatically equals the student-average effect.

Power would be determined through simulation after a separate feasibility study supplies plausible task-score variability, intraclass correlations, cluster sizes, baseline-outcome association, and attrition. Educators and assessment specialists would define the smallest educationally worthwhile improvement in advance, using its implications for competent physical performance and instructional cost. Simulations would vary uncertain design parameters and report the number of clusters required across credible scenarios. The paper supplies no fabricated sample size or power percentage. If recruitment cannot support the registered precision target, the project would remain a feasibility study and its claims would be narrowed.

The primary outcome, contrast, covariates, scoring decisions, and analysis code would be registered before assignment. H1 would constitute the sole primary confirmatory test. Secondary estimates would carry confidence intervals and an explicitly declared multiplicity policy; exploratory subgroup analyses would be labelled accordingly. Mediation through revision quality would remain exploratory because the mediator is not randomised and post-treatment confounding may remain. A model that fits a plausible pathway is insufficient evidence that the sequence operates through that pathway.

Missingness would be tabulated by assignment, cluster, access route, reason, and assessment point. Absence, withdrawal, discomfort, equipment failure, and inability to schedule a physical task would remain distinguishable. The primary missing-data approach would be specified using the feasibility study and might use multilevel multiple imputation under a stated missing-at-random assumption. Complete-case estimates and prespecified departures from that assumption would be sensitivity analyses. Discomfort-related noncompletion would also be reported as an implementation outcome, not imputed away as if it were educationally irrelevant.

The registered estimand would not condition on lesson completion, use of a headset, or later consent to publicity. Per-protocol comparisons and route-specific summaries would be secondary descriptions with their selection limitations stated. A school leaving the study would count as cluster attrition even if replacement recruitment is possible; a new school would not silently replace it in the original randomisation. All deviations, scoring changes, and analytic revisions would have a dated amendment trail.

11. Minors, research governance, and participant protection

Before recruitment, the responsible institution would obtain an appropriate independent ethics determination and resolve jurisdictional requirements, data agreements, device age restrictions, and school permissions. In settings governed by the United States HHS research regulations, Subpart D provides additional protections for children, including applicable requirements for parental permission and child assent. Whether a particular school study is exempt, minimal risk, or subject to other requirements is an institutional determination; this manuscript does not confer an exemption or approval. [R10]

Students would receive age-appropriate explanations and a practical way to decline or stop research participation without academic penalty. A conventional or accessible lesson route would remain available. A student’s refusal would override a wish to obtain a visually impressive demonstration. Staff would check comfort, provide immediate stopping options, maintain a safe physical area, and record adverse experiences using a prespecified escalation plan. The trial would not expose students to hazardous loads, high voltage, or unsupervised workshop machinery.

Identities and research data would be separated under role-based access, with a documented retention schedule and secure deletion process. Public release would prioritise task specifications, code, aggregate findings, and synthetic demonstration records. A de-identification review would precede any release of student-level data, especially where small schools, rare accommodations, or distinctive projects make reidentification plausible. Commercial platforms would not receive student research

records for model training. Photography, public authorship, and public agent demonstrations would require separate, genuinely optional permissions.

12. Decision rules and practical implementation

A district should adopt an evidential decision ladder. First establish feasibility, accessible participation, and reliable scoring. Next test the registered primary outcome with sufficient precision. Only then examine cost and staffing alongside educational benefit. A positive immediate test with no delayed transfer would support a short-term learning claim, not durable competence. A null result with a wide interval would support further uncertainty, not proof of no effect. A sufficiently precise interval excluding the prespecified worthwhile improvement would argue against scaling the present design.

Three findings would directly challenge the proposed framework. More vivid engagement without better delayed physical performance would reject the intervention’s principal educational claim. More revisions without better evidence-based explanations would challenge the revision mechanism. A benefit confined to students already comfortable with the device, accompanied by avoidable participation barriers, would challenge its inclusive policy value. Negative or mixed findings should trigger a revised curriculum or narrower use case rather than a change of outcome after the fact.

Implementation records would include staff preparation, lesson time, equipment maintenance, route support, technical interruptions, and opportunity cost. These quantities would be reported per participating class and student, with amortisation assumptions made explicit rather than hidden in a cost-per-headset figure. Schools would retain a usable lesson if a vendor changes terms or a device fails. Designership infrastructure should preserve the consequence map, versioned lesson, scoring rubric, and amendment history so that another teacher can inspect what actually produced a reported effect.

13. Boundaries, contribution, and research transparency

This protocol tests a specific instructional sequence within a limited scientific domain. It cannot establish that Movement Thinking is a general theory of cognition, that XR reorganises children’s brains, or that digital campuses replace the social functions of schools. Teacher expectations, contamination, variable access routes, imperfect masking, and selective school participation remain important threats. The inclusive policy estimand improves practical relevance while deliberately leaving the independent effect of immersion unresolved. Replication would need new tasks, sites, age groups, and accessibility conditions.

The contribution is a shift in what a school asks technology to prove. A successful environment would make consequential scientific relationships available to diverse

learners, preserve the reasoning behind revision, and support competent action after the environment is removed. Such evidence could justify a place for XR in a coherent STEM curriculum. Its absence would be equally informative. Designership becomes a research programme when attractive experiences are accompanied by credible tests of their limits.

Author and assistance disclosure: Dr. Freedom Cheteni is the intended human author and originator of the Movement Thinking orientation used in this manuscript. AI assistance supported literature discovery, synthesis, protocol drafting, and editorial development; the human author remains responsible for approving the claims, verifying the manuscript, and addressing corrections. No AI system is an author or human peer reviewer. Independent human peer review and ethics approval are not claimed. No trial data, funding award, external institutional affiliation, or registered DOI has been created by this article.

References and verification record

References retain their original authorship and publication identity. Verification dates below record source and metadata checking; they do not confer review or approval on this article.

[R1] Carly Kontra; Daniel J. Lyons; Susan M. Fischer; Sian L. Beilock (2015). Physical Experience Enhances Science Learning. Psychological Science, 26, (6), 737-749.

DOI: https://doi.org/10.1177/0956797615569355

Source: https://journals.sagepub.com/doi/10.1177/0956797615569355

Verified 2026-10-16. Publisher page and Crossref metadata agree on authors, title, journal and 2015 publication. College physics evidence; no K-12 efficacy extrapolation.

[R2] Robb Lindgren; Michael Tscholl; Shuai Wang; Emily Johnson (2016). Enhancing learning and engagement through embodied interaction within a mixed reality simulation. Computers & Education, 95, 174-187.

DOI: https://doi.org/10.1016/j.compedu.2016.01.001

Source: https://stars.library.ucf.edu/scopus2015/2783/

Verified 2026-10-16. Author-institution repository confirms complete author list/title; Crossref confirms DOI, venue, April 2016, pages 174-187. Publisher abstract identified; full publisher page fetch failed, so no detailed numerical claims used.

[R3] Mina C. Johnson-Glenberg (2018). Immersive VR and Education: Embodied Design Principles That Include Gesture and Hand Controls. Frontiers in Robotics and AI, 5.

DOI: https://doi.org/10.3389/frobt.2018.00081

Source: https://www.frontiersin.org/journals/robotics-and-ai/articles/10.3389/frobt.2018.00081/full

Verified 2026-10-16. Full publisher article and Crossref metadata checked. Design/theoretical account, not new efficacy estimate.

[R4] Guido Makransky; Thomas S. Terkildsen; Richard E. Mayer (2019). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. Learning and Instruction, 60, 225-236.

DOI: https://doi.org/10.1016/j.learninstruc.2017.12.007

Source: https://researchprofiles.ku.dk/en/publications/adding-immersive-virtual-reality-to-a-science-lab-simulation-caus/

Verified 2026-10-16. Author-institution abstract and Crossref confirm complete attribution, university sample, comparison, and 2019 issue publication. DOI string contains 2017 but article issue is 2019.

[R5] Guido Makransky; Niels K. Andreasen; Sarune Baceviciute; Richard E. Mayer (2021). Immersive virtual reality increases liking but not learning with a science simulation and generative learning strategies promote learning in immersive virtual reality. Journal of Educational Psychology, 113, (4), 719-735.

DOI: https://doi.org/10.1037/edu0000473

Source: https://researchprofiles.ku.dk/en/publications/immersive-virtual-reality-increases-liking-but-not-learning-with-/

Verified 2026-10-16. Author-institution abstract confirms high-school experiments and conditional enactment result; Crossref confirms 2021 issue 113(4), 719-735. Copyright/early-online 2020 is distinct from 2021 issue year.

[R6] Guido Makransky; Gustav B. Petersen (2021). The Cognitive Affective Model of Immersive Learning (CAMIL): a Theoretical Research-Based Model of Learning in Immersive Virtual Reality. Educational Psychology Review, 33, (3), 937-958.

DOI: https://doi.org/10.1007/s10648-020-09586-2

Source: https://link.springer.com/article/10.1007/s10648-020-09586-2

Verified 2026-10-16. Full publisher article and Crossref confirm 6 January 2021 publication, authors and theoretical status.

[R7] Susan M. Barnett; Stephen J. Ceci (2002). When and where do we apply what we learn?: A taxonomy for far transfer. Psychological Bulletin, 128, (4), 612-637.

DOI: https://doi.org/10.1037/0033-2909.128.4.612

Source: https://pubmed.ncbi.nlm.nih.gov/12081085/

Verified 2026-10-16. Crossref DOI metadata confirms complete author/title/year/venue/page range; PubMed search record independently matches title and authors. Used as transfer taxonomy, not an intervention result.

[R8] World Wide Web Consortium (W3C) (2021). XR Accessibility User Requirements. W3C Working Group Note, 25 August 2021.

Source: https://www.w3.org/TR/xaur/

Verified 2026-10-16. Official W3C full document and publication history verify title/date/Working Group Note status. Not a W3C Recommendation or product certification.

[R9] Marion K. Campbell; Gilda Piaggio; Diana R. Elbourne; Douglas G. Altman; CONSORT Group (2012). Consort 2010 statement: extension to cluster randomised trials. BMJ, 345, (sep04 1), e5661-e5661.

DOI: https://doi.org/10.1136/bmj.e5661

Source: https://www.bmj.com/content/345/bmj.e5661

Verified 2026-10-16. BMJ publisher search and Crossref confirm title, four named authors, CONSORT Group attribution, 4 September 2012 date and DOI. Cluster reporting guidance adapted; no claim that guidance alone constitutes trial registration.

[R10] United States Department of Health and Human Services, Office for Human Research Protections (2024). 45 CFR 46, Subpart D-Additional Protections for Children Involved as Subjects in Research. HHS regulations; official page records amendment 28 October 2024.

Source: https://www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/common-rule-subpart-d/index.html

Verified 2026-10-16. Official regulatory text checked, including applicability and permission/assent provisions. 2024 is the amendment shown on current page, not original enactment date (1983). Applicability/approval remain institutional determinations.

Source and authorship notes

The seed attachment supplied themes and proposed studies, not a verified bibliography, completed data, or evidence of independent human peer review.

The manuscript is an original argument and prospective design; cited papers remain attributed to their original authors and venues and are not republished here.

Literature discovery and metadata verification conducted 2026-10-16. Eight DOI references verified against Crossref plus publisher/author-institution sources; two official governance/accessibility documents verified directly.

The author-approved Movement Thinking orientation is a proposed framework. This manuscript does not claim it has already been experimentally validated.

No primary dataset has been analysed; no quantitative effect size, achieved power, participant sample, ethics approval, DOI registration, or completed human peer review is claimed.

The proposed trial is not prospectively registered yet. Registration and independent ethics determination must precede recruitment.

D02 / Human pedagogical authority

Who Holds the Classroom? Bounded Pedagogical Autonomy for Social Robots and AI Agents in K-12 Education

Dr. Freedom Cheteni

The Research University | Volume 1, 2026 | Edition 1.0 | D02

Critical integrative analysis and prospective field-study protocol

Original analyses and prospective protocols. New independent human peer review is not recorded for this edition.

DOI: not assigned to this research edition.

Read online | Author ORCID: 0009-0006-3507-0094 | Editorial record

Abstract

Social robots and generative AI agents create a consequential distinction between executing instructional routines and holding pedagogical authority. This original critical integrative analysis proposes bounded pedagogical autonomy: action-specific permissions that preserve human responsibility for curriculum, interpretation, safeguarding, and consequential decisions. Primary classroom and laboratory studies are examined for the conditions under which personalisation, embodiment, rapport, and social influence may support or undermine learning; findings from language-learning and generative-AI trials are interpreted within their actual populations and assessment conditions. The article introduces an autonomy-oversight matrix and a six-part failure taxonomy, then specifies a prospective four-arm school-cluster randomised trial crossing embodied versus screen presentation with adaptive versus fixed scaffolding. The primary outcome is delayed unaided transfer rather than assisted performance or engagement. The protocol matches instructional content and adult contact, measures workload and cost, specifies clustered analysis and missing-data sensitivity, and makes consent, child assent, accessibility, escalation, and data minimisation operational requirements. Recruitment, ethics approval, preregistration, power simulation, and trial execution have not occurred. The contribution is a falsifiable design and governance argument, not a report of efficacy. Designership is advanced as accountable inquiry: deployment must remain contingent on independent learning, sustainable adult oversight, and protection of children’s rights.

Keywords: Designership, social robots, K-12 education, pedagogical authority, bounded autonomy, unaided transfer, child-centred AI, cluster randomised trial

Why classroom autonomy needs a theory of authority

A social robot can ask a question, wait for an answer, and invite a child to revise an explanation. None of those actions establishes that the robot should decide what the child is entitled to learn, whether an answer deserves a grade, or when a family should be contacted. The important educational distinction is between executing an instructional routine and holding pedagogical authority. This article develops bounded pedagogical autonomy: a design in which an agent may select a limited next action, while human educators retain responsibility for curriculum, interpretation, safeguarding, and consequential decisions.

The intellectual starting point is the Designership programme’s interest in social robotics, agent-supported classrooms, and Movement Thinking. These are treated as a research agenda rather than evidence of effectiveness. The article makes three original contributions: an autonomy-oversight matrix that connects permissions to consequences; a failure taxonomy connecting technical incidents to learning and dignity; and a comparative field-study protocol capable of separating embodied presence from adaptive pedagogy. It is a critical integrative analysis and prospective protocol, not a completed trial, systematic review, or validated instrument.

The central proposition is deliberately conditional. An agent is educationally useful when it helps a learner produce a better explanation and subsequently act without the agent, while leaving educators able to understand and contest its interventions. A fluent exchange, longer screen time, or apparent affection for a robot is insufficient. For Designership, the consequential question is therefore not whether a classroom can run itself. It is which bounded forms of automation make inquiry more capable, equitable, and accountable.

Evidence is conditional, not a verdict on robots

Classroom robotics research provides reasons for investigation and reasons for restraint. Baxter and colleagues studied personalised and non-personalised robot peers in two primary-school classrooms over two weeks. They reported learning advantages associated with personalisation, including on a novel topic. The small number of classrooms and the specific robot-touchscreen task limit generalisation to whole-school deployment. Importantly, operation without experimenter supervision did not mean absence of the classroom teacher. The study supports testing particular personalised routines, not substituting machines for adult educational responsibility. [R1]

Tanaka and Matsuzoe investigated children teaching vocabulary to a care-receiving robot. Their field experiments make an important conceptual move: a robot can be an imperfect recipient of an explanation rather than an authoritative examiner. This motivates a classroom role in which students must explain and correct. It does not

establish that every simulated mistake improves learning; teachers must distinguish a purposeful misconception from an unmarked factual error. [R2]

Findings also challenge an assumption that more social behaviour necessarily produces better instruction. Kennedy and colleagues experimentally examined how a tutor robot's social behaviour could interfere with child learning. In a larger language-learning programme, van den Berghe and colleagues reported heterogeneous relationships between children's language-learning characteristics and robot support; the underlying comparison found no overall vocabulary advantage from adding the robot to the tablet lessons. Embodiment, gestures, instructional content, and learner characteristics should therefore be separated analytically rather than bundled as a uniformly effective technology. [R3, R4]

Social effects can be educationally ambiguous. Kory-Westlund and Breazeal's storytelling experiment examined rapport-related behaviours and found changes in children's interaction, retelling, and responses to the robot. Vollmer and colleagues showed that children could conform to incorrect robot-group answers in an experimental task. Neither study establishes inevitable classroom harm. Together they justify measuring whether a learner can challenge a persuasive agent, rather than treating compliance or relationship ratings as an unqualified benefit. [R5, R6]

What generative AI changes-and what it does not

Language models broaden the range of responses an agent can produce, but they also make its instructional behaviour less exhaustively scriptable. A robot that selects among approved prompts and a robot that improvises explanations are different interventions. The second requires monitoring the semantic content of outputs, not merely testing whether motors and microphones work. Retrieval from approved material can help constrain responses, but it does not guarantee correct inference, faithful quotation, or an appropriate response to a child.

A high-school mathematics trial by Bastani and colleagues is especially relevant because it distinguished assisted task performance from subsequent unaided performance and compared different forms of AI support. Its results warn against equating access to an effective answer generator with durable learning. The publication has a linked correction, which is retained in this article's evidence record; it corrects Osbert Bastani's institutional affiliation following a production error, rather than the study findings, and is not treated as an independent efficacy study. No effect size is transferred from that intervention to classroom robots. [R7, R8]

Kestin and colleagues reported beneficial learning outcomes from a carefully designed AI tutor compared with in-class active learning in college physics. That study concerns university students and a particular instructional design. Its appropriate role here is to show that implementation and pedagogy matter, not to predict benefits for children. Together, these studies support a test of competing mechanisms: scaffolding may

promote independent reasoning, whereas immediate answer access may substitute for the reasoning that assessment intends to reveal. [R9]

The proposed system therefore offers prompts for explanation, comparison, and revision rather than unrestricted answer delivery. The lesson's explanation of the tool is literal: it is software that can be useful and wrong, and adults remain responsible for teaching. The interface must avoid claims that the agent understands a child's feelings, has needs the child must satisfy, or is a confidential substitute for a trusted adult.

The autonomy-oversight matrix

The proposed matrix classifies an action along two axes: its educational consequence and the reversibility of its effects. Consequence is not identical to technical complexity. Reading an approved sentence aloud is technically sophisticated but relatively low consequence; silently assigning a child to a remedial pathway can be technically simple and highly consequential. Reversibility concerns whether a mistake can be corrected without persistent educational, social, or informational damage. A permission is granted to a specified action in a specified context, never to an agent in general.

Level A permits bounded execution: announcing the next approved task, reading a source passage, or offering a preapproved prompt. The teacher approves the lesson and prompt set beforehand; the system logs the action and offers an immediate stop control. Level B permits constrained adaptation: selecting among approved hints using the learner's current task response. The teacher can inspect the basis for the selection, override it, and review a concise exception log. The agent cannot infer a general ability label from that interaction.

Level C covers proposals requiring explicit adult approval before their effect: changing a learning goal, recommending a placement, sending an external message, or adding information to a persistent student record. The agent may draft a proposal with its evidence and uncertainty; it cannot execute it. Level D excludes delegated authority in this study: discipline, admission, disability identification, mental-health diagnosis, high-stakes grading, and independent safeguarding decisions. A child's disclosure of possible harm triggers an adult safeguarding pathway; it does not trigger an automated clinical judgement.

These levels become an implementation table with five mandatory fields for each action: permitted input, permitted output, responsible adult, escalation trigger, and maximum retention period. For example, an uncertain oral response permits a repeat-or-type request, not an ability score. A request for a grade permits referral to the teacher, not a generated mark. A request to share personal information permits a privacy reminder and adult referral, not a new data collection form. The table is an original governance proposal that the trial will evaluate, not an established standard.

A failure taxonomy that preserves educational meaning

A failure register should distinguish six classes. Epistemic failures include incorrect explanations, invented sources, and misleading certainty. Pedagogical failures include revealing an answer too early, rewarding an unsupported claim, or giving a hint that makes the assessed task trivial. Interaction failures include speech-recognition breakdown, inaccessible instructions, and turn-taking that prevents a learner from participating. Authority failures occur when the agent acts outside its permission level or obscures who made a consequential decision. Relational failures include coercive praise, pressure to agree, or language encouraging secrecy. Data failures include excessive collection, wrong-student attribution, and unauthorised disclosure.

The unit of coding is a specific event with an observable context and consequence, not a general impression that a session went poorly. Independent trained coders will record the task, model and prompt version, output, permission level, severity, student-visible correction, time to adult response, and whether the learning activity resumed. A wrong statement caught before delivery differs from one repeated by a learner on an unaided transfer task. Both matter, but they should not be counted as equivalent instructional exposure.

Severity rules distinguish inconvenience, recoverable instructional distortion, and serious safety or rights events. All serious events trigger immediate suspension of the affected function pending human review. A single credible unauthorised disclosure or safeguarding mishandling is sufficient for that suspension; there is no requirement to wait for statistical significance. For ordinary errors, a preregistered review threshold will combine exposure frequency and severity. Thresholds will be set with educators, family representatives, and the ethics committee before recruitment, then frozen for the trial.

Reporting includes the denominator of agent opportunities and learner exposure time. A system used less often may show fewer incidents while being less reliable per interaction. Conversely, an increase in logged incidents after better monitoring may reflect better detection. The trial will therefore report incident rates, total exposure, detection route, correction quality, and unresolved cases separately.

Field-study design: separating embodiment from pedagogy

The proposed primary study is a four-arm, parallel, school-cluster randomised trial with a two-by-two factorial design. One factor is interface: a stationary social robot or a screen-based agent with matched audio and visual content. The other is instructional policy: bounded adaptive scaffolding or a fixed, curriculum-aligned prompt sequence. All arms receive the same learning goals, source material, task time, teacher training, escalation procedure, and human supervision. This isolates the incremental contribution of embodied presentation and of adaptive prompt selection more cleanly than comparing a robot lesson with unrelated business-as-usual instruction.

Schools, rather than individual children within the same room, will be allocated to reduce intervention sharing and inconsistent classroom norms. Schools will be recruited before allocation; classes and consented students will be enumerated before randomisation where feasible. An independent statistician will use concealed, blocked allocation stratified by school phase, baseline attainment band, and broad deprivation category. The analysis will respect the unit of assignment and report the number of schools, classes, teachers, and students in every arm. Cluster-trial reporting principles inform this design. [R10]

The initial target population is students aged approximately ten to thirteen undertaking an eight-week science explanation and design unit. The narrow age range and subject context are deliberate boundaries, not a claim to cover all K-12 education. Tasks require learners to explain a phenomenon, compare evidence, revise a proposed design, and justify a trade-off. Teachers may stop a session at any time. Students can choose an equivalent non-agent activity without academic penalty, with participation in research data collection handled separately from access to instruction.

A two-week familiarisation period precedes the eight-week instructional unit. All groups receive comparable orientation to their assigned interface and equivalent researcher contact. The robot remains stationary and does not touch children or navigate around a classroom. A neutral introduction and matched interaction time reduce novelty differences, but cannot eliminate them. Weekly novelty ratings and exposure histories will therefore be reported as process measures rather than used to redefine the primary outcome after results are known.

Outcomes that require the learner to act independently

The primary endpoint is performance on a delayed, unaided transfer assessment four weeks after the unit. Students encounter unfamiliar surface features but the same underlying scientific and design principles. They produce an explanation and revise a design in response to new evidence without access to any agent. A prespecified rubric scores causal adequacy, use of evidence, recognition of a trade-off, and justified revision. Two raters masked to treatment will independently score a prespecified sample sufficient to estimate agreement; discrepancies follow a documented adjudication rule.

Before recruitment, assessment developers will pilot task difficulty, accessibility, rubric interpretation, and parallel-form comparability with students outside trial schools. The primary score is a fixed weighted composite established during that pilot. It will not be reweighted to favour whichever subscale yields the largest intervention difference. Immediate knowledge, retention, and revision quality are secondary outcomes. Engagement is a process measure. The capacity to detect and challenge a plausible but incorrect explanation is a separately declared secondary outcome using a brief, debriefed assessment rather than covert manipulation during instruction.

Teacher workload is a coequal implementation criterion. Weekly short diaries will record preparation, supervision, technical recovery, reviewing exceptions, and additional time outside paid hours. A subsample will complete structured time observations to estimate diary bias. Costs will include devices, software, accessibility supports, training, maintenance, and adult attention. The study will report educational effects alongside workload and cost rather than converting all benefits into a single promotional return-on-investment figure.

Accessibility will be designed into the tasks, including text alternatives, captions, adjustable response time, switch-compatible input where required, and a route that does not require speech. Language background and disability-related accommodation are not error labels. Recruitment and retention by relevant groups will be reported with privacy-preserving cell rules. Interpretation of subgroup differences will attend to whether the interface offered equal access before attributing differences to learners.

Hypotheses, estimands, and a power plan

The primary confirmatory hypothesis is that bounded adaptive scaffolding improves delayed unaided transfer relative to the fixed prompt sequence, averaged across interface conditions. The primary estimand is the student-average intention-to-treat difference for students eligible and enrolled under the preregistered recruitment rule, accounting for school assignment. Marginal contrasts will weight enrolled students equally; school-size weights will be fixed using preallocation eligible enrolment, and the analysis plan will specify their implementation and handling of informative cluster size before allocation. An equal-school-weight contrast will be a sensitivity analysis because it targets a different population quantity. A second confirmatory contrast tests the main effect of embodiment on that same outcome. The family of these two tests will use a prespecified multiplicity procedure. The interface-by-policy interaction is secondary unless power simulations support an adequately sized confirmatory interaction test before recruitment begins.

An original mechanism hypothesis predicts that adaptive scaffolding increases substantive explanation revisions rather than merely response length. A rival account predicts improved assisted task completion without improved unaided transfer. A separate boundary hypothesis predicts that any learning benefit is unacceptable for deployment if accompanied by a meaningful increase in uncompensated teacher workload or a serious rights event. The workload margin must be negotiated and documented with teachers before allocation; this manuscript does not invent a universally acceptable number of extra minutes.

Power will be established through simulation using the planned multilevel analysis, plausible school and class intraclass correlations, unequal cluster sizes, baseline-outcome correlation, and expected attrition. A provisional planning scenario is sixty schools, fifteen per factorial cell, with approximately forty participating students per school before attrition. This is a feasibility scenario, not a recruited sample or a power finding. Simulations will examine small educationally relevant effects across conservative correlation and attrition scenarios; the final sample target, code, assumptions, and resulting detectable effects will be registered before randomisation.

If the feasible number of schools cannot support the declared estimands, investigators will increase recruitment, reduce the number of confirmatory contrasts, or designate the study as a feasibility trial before allocation. They will not justify an underpowered efficacy claim by counting thousands of student interactions as independent observations. School withdrawal will be simulated separately from student nonresponse because losing clusters damages information differently from losing the same number of students across retained schools.

Analysis, missingness, and evidence of mechanism

The primary analysis will use a hierarchical model with treatment indicators, their prespecified interaction, baseline transfer-related attainment, randomisation strata, and school and class structure appropriate to the sampling plan. Small-sample degrees-of-freedom corrections or randomisation-based inference will be prespecified for the number of clusters. Estimates will include confidence intervals and both raw-score and standardised interpretations. A decision to use random rather than fixed class effects will be made before unmasking, based on the data structure and simulation performance, not statistical significance.

Missing outcomes will be tabulated by school, arm, time point, and stated reason. The primary missing-data approach will use a multilevel-compatible procedure under an explicit missing-at-random assumption, including baseline attainment and predictors of nonresponse. Sensitivity analyses will vary departures from that assumption, including worse unobserved outcomes among disengaged students and informative school withdrawal. A complete-case estimate will be reported as a sensitivity analysis, not treated as automatically unbiased. Attendance, tool exposure, crossover, and teacher overrides will be reported without conditioning the primary treatment effect on post-randomisation compliance.

Potential mediation by revision quality will be exploratory. Random assignment to a tool does not randomise a student’s willingness to revise; a simple association between revision and transfer cannot establish that revision caused the treatment effect. The study will analyse temporal patterns and contrasting cases to refine the proposed mechanism, explicitly acknowledging confounding and measurement error. Interview accounts from teachers and students will be used to understand failures and implementation conditions, not to replace a null primary outcome with a success narrative.

All prespecified outcomes, model versions, subgroup analyses, and deviations will appear in a public results register. Subgroups will be limited to a small set justified before allocation, with interaction estimates and uncertainty. Any additional discovery analysis will be labelled exploratory and will require later replication. A null embodiment effect remains informative: if a less expensive screen produces comparable transfer with fewer burdens, the appropriate educational conclusion may favour that screen.

Failure to detect an interface effect will not establish equivalence. An equivalence or non-inferiority claim would require a defensible educational margin, appropriate power, and a prespecified testing procedure before recruitment. The present protocol is a superiority design. A confidence interval that excludes only large differences may narrow a deployment decision, but it cannot certify that two tools are educationally interchangeable. Any later equivalence study will be registered as a distinct analysis or trial rather than retrofitted to a null result.

Consent, safeguarding, and data minimisation

The study requires approval from an appropriately constituted research ethics body, participating schools, and relevant data controllers before recruitment. Guardian permission and age-appropriate child assent will be obtained for research participation; children can pause or withdraw without losing access to equivalent learning. Consent materials will explain what the agent does, who can inspect records, what data leave the device, and the limits of confidentiality when an adult must respond to a safeguarding concern. No approval, consent, or recruitment is claimed in this article.

Default data collection excludes facial emotion inference, continuous classroom video, biometric identification, and ambient recording of nonparticipants. Speech is optional; push-to-talk input will be transient where technically feasible, with text retained only when necessary for the declared outcomes and incident audit. Research identifiers are separate from school identifiers, access is role-limited, and retention and deletion schedules are approved before data collection. Student data will not be used to train general-purpose models or support advertising. These are proposed operational safeguards, to be tested through technical and institutional review rather than asserted as automatic compliance. [R12, R14]

The interface offers a visible stop-and-ask-an-adult control. Disclosure of possible harm causes the agent to cease the ordinary tutoring routine, provide a short neutral acknowledgement, and alert the designated adult through the school’s approved route. It must not promise secrecy, investigate an allegation, or make a diagnosis. Escalation includes a documented fallback when the designated adult is unavailable. A practical drill with adults and simulated inputs must pass before any child-facing session.

Relational restraint is part of safeguarding. The agent will not ask children to keep it company, express distress when they leave, solicit personal secrets, or imply that declining interaction is unkind. A student may question its accuracy without losing rewards. These restrictions may reduce some measures of attachment or compliance; that is compatible with the educational objective. UNICEF’s child-rights guidance supports attention to safety, privacy, fairness, participation, and accountable oversight, but does not itself validate this protocol or endorse the system. [R12]

Reproducibility and a publication record fit for scrutiny

Before the trial, investigators will preregister the protocol, outcome definitions, analysis plan, failure codebook, stopping rules, and intervention manifest in an appropriate public registry. The manifest will identify robot hardware, operating software, language model provider and version, system instructions, retrieval corpus, allowed actions, and input/output filters. Material system changes during the trial require a dated amendment and a plan for interpreting affected observations. Frozen instructional behaviour is preferable to silently adopting a provider update that changes the intervention halfway through the study.

The SPIRIT-AI and CONSORT-AI guidance was developed for clinical research, not schools; its requirements concerning system versions, inputs, outputs, human interaction, and error analysis are adapted here as reporting aids. Their use does not turn an educational trial into a clinical trial or establish institutional approval. Together with the cluster-trial extension, they motivate a record that allows readers to identify which intervention was actually evaluated and under what human operating conditions. [R10, R11, R13]

Publication will include assessment rubrics, analysis code, simulated or synthetic test data, aggregated incident tables, and a description of restricted data access where consent and disclosure risk preclude open student-level records. Student work will be published only under separate, informed permission and an appropriate licence. Raw child conversations are not a necessary public supplement. Open research should make claims inspectable without making children permanently searchable.

The paper agent attached to this article should answer from this manuscript and its verified sources, distinguish proposed methods from completed findings, and expose the passage supporting a response. It should return uncertainty when the manuscript is silent. Successful retrieval is not evidence that the field trial has occurred, and software evaluation must be recorded separately from scholarly peer review. Independent human reviewers should examine the theoretical argument, child protections, outcome validity, and analysis plan before an efficacy study begins.

Limitations and a decision rule for Designership

The framework is an original proposal awaiting validation. Its matrix may be too coarse for some classrooms, and coding failures reliably may require more training than expected. School recruitment may favour well-resourced settings, limiting transportability. Students and teachers cannot be blinded to a physical robot, and their expectations may influence outcomes despite matched contact and assessment masking. A short instructional unit cannot establish long-term effects on attachment, teacher employment, civic development, or whole-school culture.

The factorial design estimates embodiment within a tightly specified lesson system; it does not compare every robot with every chatbot. The robot and screen cannot be made sensorially identical, and costs may differ substantially. Teacher adaptations may be both necessary for inclusion and a source of implementation variation. Results must therefore report what educators actually did, including overrides and departures, rather than attributing all change to an abstract category called AI.

A defensible deployment decision requires a joint pattern: credible evidence for the stated claim about unaided learning, no unacceptable rights or safety events, workable adult oversight, accessible participation, and sustainable workload and cost. Positive engagement with weak transfer fails that test. Improved transfer produced by hidden teacher labour also fails the claim that automation has increased classroom capacity. The study may support a narrower permission set, a screen rather than a robot, or no adoption. Designing the inquiry so that each conclusion remains possible is the article's central contribution to Designership.

References and verification record

References retain their original authorship and publication identity. Verification dates below record source and metadata checking; they do not confer review or approval on this article.

[R1] Paul Baxter; Emily Ashurst; Robin Read; James Kennedy; Tony Belpaeme (2017). Robot education peers in a situated primary school study: Personalisation promotes child learning. PLOS ONE.

DOI: https://doi.org/10.1371/journal.pone.0178126

Source: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0178126

Verified 2026-10-16. Primary classroom field experiment. Publisher full text and Crossref DOI metadata verified. Two classrooms; personalisation evidence is not whole-school autonomy evidence.

[R2] Fumihide Tanaka; Shizuko Matsuzoe (2012). Children Teach a Care-Receiving Robot to Promote Their Learning: Field Experiments in a Classroom for Vocabulary Learning. Journal of Human-Robot Interaction.

DOI: https://doi.org/10.5898/jhri.1.1.tanaka

Source: https://doi.org/10.5898/JHRI.1.1.Tanaka

Verified 2026-10-16. Primary field experiments. Publisher DOI landing, University of Tsukuba laboratory bibliography and Crossref metadata verified; no unrestricted classroom-autonomy claim.

[R3] James Kennedy; Paul Baxter; Tony Belpaeme (2015). The Robot Who Tried Too Hard: Social Behaviour of a Robot Tutor Can Negatively Affect Child Learning. Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction.

DOI: https://doi.org/10.1145/2696454.2696457

Source: https://doi.org/10.1145/2696454.2696457

Verified 2026-10-16. Primary conference experiment, not journal article. Crossref title/subtitle and author order verified; official HRI programme and author institutional bibliography confirm study identity. DOI publisher full text blocked to web retrieval; cited conclusion corroborated by primary author record.

[R4] Rianne van den Berghe; Ora Oudgenoeg-Paz; Josje Verhagen; Susanne Brouwer; Mirjam de Haas; Jan de Wit; Bram Willemsen; Paul Vogt; Emiel Krahmer; Paul Leseman (2021). Individual Differences in Children's (Language) Learning Skills Moderate Effects of Robot-Assisted Second

Language Learning. Frontiers in Robotics and AI.

DOI: https://doi.org/10.3389/frobt.2021.676248

Source: https://www.frontiersin.org/journals/robotics-and-ai/articles/10.3389/frobt.2021.676248/full

Verified 2026-10-16. Primary study and follow-up analysis from the same L2TOR dataset, not an independent replication of Vogt et al. Publisher full text verifies authors, null overall added robot benefit and mixed moderator results.

[R5] Jacqueline M. Kory-Westlund; Cynthia Breazeal (2019). Exploring the Effects of a Social Robot's Speech Entrainment and Backstory on Young Children's Emotion, Rapport, Relationship, and Learning. Frontiers in Robotics and AI.

DOI: https://doi.org/10.3389/frobt.2019.00054

Source: https://www.frontiersin.org/journals/robotics-and-ai/articles/10.3389/frobt.2019.00054/full

Verified 2026-10-16. Primary experiment. Publisher full text verifies two authors, 86 children and 2x2 design; no evidence of established developmental benefit inferred from rapport.

[R6] Anna-Lisa Vollmer; Robin Read; Dries Trippas; Tony Belpaeme (2018). Children conform, adults resist: A robot group induced peer pressure on normative social conformity. Science Robotics.

DOI: https://doi.org/10.1126/scirobotics.aat7111

Source: https://www.science.org/doi/10.1126/scirobotics.aat7111

Verified 2026-10-16. Primary experimental study. Science Robotics DOI and Crossref author order/title/year verified; publisher search and author institutional record corroborate child conformity finding. Publication day not asserted because indexes differ.

[R7] Hamsa Bastani; Osbert Bastani; Alp Sungu; Haosen Ge; Özge Kabakcı; Rei Mariman (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. Proceedings of the National Academy of Sciences.

DOI: https://doi.org/10.1073/pnas.2422633122

Source: https://www.pnas.org/doi/10.1073/pnas.2422633122

Verified 2026-10-16. Primary high-school randomised trial, journal version of record. Exact DOI verified against publisher and Crossref; source-version preprint 10.2139/ssrn.4895486 is not treated as a second study. Linked correction retained as R8.

[R8] (2025). Correction for Bastani et al., Generative AI without guardrails can harm learning: Evidence from high school mathematics. Proceedings of the National Academy of Sciences.

DOI: https://doi.org/10.1073/pnas.2518204122

Source: https://www.pnas.org/doi/10.1073/pnas.2518204122

Verified 2026-10-16. Publisher correction notice, not a study. Crossref update-to explicitly identifies correction to R7; primary full text independently checked in editorial audit: it corrects Osbert Bastani’s institutional affiliation after a production error, not the findings. Crossref notice carries no individual authors. Complete ongoing retraction surveillance is not claimed.

[R9] Greg Kestin; Kelly Miller; Anna Klales; Timothy Milbourne; Gregorio Ponti (2025). AI tutoring outperforms in-class active learning: an RCT introducing a novel research-based design in an authentic educational setting. Scientific Reports.

DOI: https://doi.org/10.1038/s41598-025-97652-6

Source: https://www.nature.com/articles/s41598-025-97652-6

Verified 2026-10-16. Primary randomised trial in college physics, not K12 or embodied robotics. Publisher full text and DOI metadata verified. Generalisation to children explicitly limited.

[R10] Marion K. Campbell; Gilda Piaggio; Diana R. Elbourne; Douglas G. Altman; CONSORT Group (2012). Consort 2010 statement: extension to cluster randomised trials. BMJ.

DOI: https://doi.org/10.1136/bmj.e5661

Source: https://www.bmj.com/content/345/bmj.e5661

Verified 2026-10-16. Primary reporting guideline, cluster-trial extension. Publisher/author metadata and CONSORT Group attribution verified. Used as reporting and design guidance, not educational efficacy evidence.

[R11] Xiaoxuan Liu; Samantha Cruz Rivera; David Moher; Melanie J. Calvert; Alastair K. Denniston; The SPIRIT-AI and CONSORT-AI Working Group (2020). Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nature Medicine.

DOI: https://doi.org/10.1038/s41591-020-1034-x

Source: https://www.nature.com/articles/s41591-020-1034-x

Verified 2026-10-16. Primary consensus reporting guideline. Publisher confirms five lead authors plus named working group. Clinical origin explicitly stated; adaptation to educational research is proposed, not validated.

[R12] UNICEF Innocenti - Global Office of Research and Foresight (2025). UNICEF Guidance on AI and Children 3.0. UNICEF Innocenti.

Source: https://www.unicef.org/innocenti/reports/policy-guidance-ai-children

Verified 2026-10-16. Authoritative primary policy guidance, version 3.0 December 2025. Official page and linked report verify child-rights requirements. No DOI assigned in consulted source; not an efficacy study or endorsement.

[R13] Samantha Cruz Rivera; Xiaoxuan Liu; An-Wen Chan; Alastair K. Denniston; Melanie J. Calvert; The SPIRIT-AI and CONSORT-AI Working Group (2020). Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. Nature Medicine.

DOI: https://doi.org/10.1038/s41591-020-1037-7

Source: https://www.nature.com/articles/s41591-020-1037-7

Verified 2026-10-16. Primary consensus protocol guideline. Publisher and PubMed verify five lead authors and working group. Crossref request rate-limited; publisher DOI/title/authors/year independently accessible. Clinical origin explicitly stated.

[R14] Fengchun Miao; Wayne Holmes (2023). Guidance for generative AI in education and research. UNESCO.

DOI: https://doi.org/10.54675/EWZM9535

Source: https://unesdoc.unesco.org/ark:/48223/pf0000386693

Verified 2026-10-16. Primary UNESCO guidance. Official UNESCO record and publication verify authors, year, title and DOI; used for human agency/data governance principles, not intervention efficacy.

Source and authorship notes

Seed attachment Pasted markdown(20261016-073609).md, library file libfile_c8df81e3d25c819186a2df6c6d4dd2ce: social robotics and autonomous-classroom themes treated as commissioning ideas, not independently verified historical or efficacy evidence.

No claim is made that Cheteni founded the first autonomous classroom, established a particular Jibo intervention, or produced empirical findings; these claims were not necessary for this article and were not verified here.

Original synthesis, governance matrix, failure taxonomy and prospective protocol. No new participant data or completed intervention analysis. Not a systematic review; search completeness and quantitative synthesis are not claimed.

AI assistance supported source discovery, metadata auditing and manuscript development. Human author accountability, independent handling and human peer review must be recorded separately; no independent human review or acceptance is claimed by this draft.

References include primary experiments, a conference paper, methodological consensus guidelines and authoritative policy guidance. Citation verification establishes identity and stated claim scope, not replication or proof of effectiveness.

Fourteen references checked 2026-10-16. R7 has a linked publisher correction R8. The citation audit is not a guarantee of complete future correction/retraction surveillance.

D03 / Mentored inquiry

Public Contribution Without Prestige Proxies: Designership as a Mentored Inquiry Pathway in K-12 Education

Dr. Freedom Cheteni

The Research University | Volume 1, 2026 | Edition 1.0 | D03

Integrative conceptual analysis and prospective evaluation protocol

Original analyses and prospective protocols. New independent human peer review is not recorded for this edition.

DOI: not assigned to this research edition.

Read online | Author ORCID: 0009-0006-3507-0094 | Editorial record

Abstract

School-linked publication can make student work visible while obscuring who performed the reasoning and whether learning transfers beyond a coached artifact. This article develops public contribution without prestige proxies as a proposed framework for Designership: a mentored inquiry pathway connecting consequential questions, bounded student authority, documented revision, contributor accountability and reciprocal public value. A selective integrative analysis of original inquiry, mentorship, feedback and transfer studies motivates the framework without treating it as validated. The article specifies a prospective, cohort-cluster randomized evaluation comparing a structured sixteen-week pathway with an attention- and resource-matched inquiry opportunity. The primary estimand is the intention-to-treat difference on an independently completed unfamiliar-task reasoning-and-revision assessment, rated without author, school or venue identities. Authentic project quality, belonging, contribution explanations, persistence, cost and burden remain secondary or descriptive outcomes. The protocol addresses allocation concealment, clustered precision, attrition, fidelity, mentor conflicts, accessible assessment, child assent, guardian permission and separate consent for public dissemination. No participants have been enrolled and no empirical effects, ethics approval, registration or independent human peer-review outcome are claimed. The contribution is a falsifiable educational and editorial architecture in which publication supports learning without becoming its proxy.

Keywords: Designership, K-12 research, mentorship, student agency, authentic assessment, transfer of learning, research integrity, equity

1. The educational problem: contribution before credential

A research pathway can produce a polished article while leaving the student's reasoning obscure. Adults can select the question, repair the analysis, supply the language and arrange the venue. The resulting object may be valuable, yet its existence does not establish what the learner can independently investigate. Conversely, an investigation that identifies a failed assumption, corrects a measurement or responsibly concludes that evidence is insufficient may demonstrate substantial learning without generating an impressive publication. A journal linked to schooling therefore needs two separate warrants: a warrant for the public claim and a warrant for the student's intellectual contribution. Neither can substitute for the other.

This article proposes Designership as a mentored inquiry pathway connecting consequential questions, disciplined investigation, documented revision and accountable public contribution. Within Cheteni's proposed SIDE architecture-Scholarship, Internship, Designership and Entrepreneurship-Designership names responsibility for making and improving a defensible contribution. Movement Thinking supplies a proposed orientation toward action, consequence and revision. These terms organize the present argument; their use does not establish a validated psychological construct, a proven intervention or a historical account of program effectiveness. The empirical contribution here is a prospective test specification, not a report of participant outcomes.

The central proposition is public contribution without prestige proxies. Students should be assessed on how they frame a question, justify a method, engage counterevidence and revise a warranted claim. Institutional branding, mentor seniority, publication count, downloads and entrepreneurial recognition can describe opportunities or dissemination, but they do not measure those capacities. Designership should earn educational significance by improving the learner's performance on a new inquiry problem under comparable support, while making the origins and limits of the public work inspectable. That claim can fail even when the programme attracts attention.

2. What existing evidence warrants-and what it leaves open

The argument draws selectively on original learning and mentorship studies and on conceptual and methodological sources. It is an integrative analysis, not an exhaustive or systematic review: no claim is made that a registered search, dual screening or complete study universe has been established. Reference metadata were checked against publisher, institutional or registration records on 16 October 2026. The sources motivate design decisions; their outcomes are not combined into a synthetic effect estimate.

Barron and colleagues examined problem- and project-based learning organized around principles including appropriate learning goals, scaffolds, revision opportunities and supportive social arrangements. Their work supports designing the conditions around a project rather than equating activity with understanding [R1]. Chinn and Malhotra's conceptual analysis distinguishes epistemologically substantial inquiry from simplified classroom procedures, directing attention to how questions, evidence and explanations are generated [R14]. Together these sources motivate a pathway in which methodological decisions remain visible, without establishing that every student-authored paper constitutes authentic inquiry.

A cluster-randomized study by Taylor and colleagues tested a high-school biology module combining student investigations and online scientist mentoring; its reported achievement and attitude effects concern that combined intervention, not mentoring isolated from curriculum [R2]. Puslednik and Brennan studied an Australian rural authentic-science programme, while Edry, Sasson and Dori examined secondary research apprenticeships and mentor-student interaction [R3] [R4]. Those settings make the pathway educationally plausible. They do not resolve selection into intensive programmes or establish benefits across subjects, ages and resource conditions. The present protocol therefore compares two supported pathways rather than an enriched experience against absence of opportunity.

Schwartz and Martin's statistics studies assessed preparation for subsequent learning; Kapur's productive-failure study examined how unsuccessful initial problem solving could contribute to later transfer [R5] [R6]. These findings motivate a new-task assessment with an opportunity to learn and revise, not a general prescription to withhold instruction. Cho and MacArthur studied revision with peer and expert feedback [R7]. Yeager and colleagues investigated how adolescents interpreted critical feedback across a racial divide, while Herrera and colleagues evaluated a school mentoring programme through random assignment [R8] [R9]. None validates Designership as a package. Each identifies a condition-feedback, relational trust, structured experience or transfer-that a defensible evaluation must distinguish from publication itself.

3. The public-contribution framework

The proposed framework separates four records that are often collapsed into a single portfolio. The inquiry record shows the question, alternatives, methods, observations and changes of interpretation. The contribution record identifies who performed substantive tasks and how assistance changed the work. The learning record contains independently assessed performances under documented conditions. The public record presents the claim, supporting material, limitations and version history at a level appropriate for its audience. A student may possess a strong learning record and elect not to make a public record. This is a legitimate completion pathway rather than a lesser educational outcome.

Five commitments connect these records. First, consequentiality requires identifying an affected audience and a decision or uncertainty that could benefit from investigation. It does not require a dramatic invention. Second, bounded authority gives students real decisions over an achievable question, method or design choice while adults retain responsibility for safety and evidential standards. Third, revision must preserve a meaningful change and its reason; counting saved files is insufficient. Fourth, contestability requires an opportunity for someone other than the project sponsor to question the claim. Fifth, reciprocity asks whether the work returns a useful explanation, artifact or finding to participants or communities without treating them as promotional material.

A Design Realization Garage is proposed here as an organizational arrangement for these commitments, not a costly room specification. Its minimum components are protected inquiry time, access to suitable materials or lawful data, a mentor compact, a revision ledger and an editorial route. A library table with accessible equipment may implement these functions more faithfully than an immersive facility optimized for spectacle. Physical and digital environments can both support the arrangement; technology is justified only when it expands investigation or access. Such functional specification makes it possible to compare implementations across schools without assuming that a branded space causes learning.

The framework also distinguishes an educational audience from a scholarly venue. A local environmental memo, an accessible software demonstration, a careful replication or a data-quality report can constitute public contribution. Journal publication is one route when methods, permissions and claims justify it. It should not be guaranteed as a course reward or purchased as a credential. A negative finding and a decision not to disseminate sensitive information can satisfy the framework when the reasoning is strong. This prevents the platform's growth objectives from defining what counts as worthwhile student work.

4. Mentoring as accountable scaffolding

Mentors require a compact describing permissible help and the decisions reserved for learners. They may model how to inspect a source, ask students to consider an alternative explanation, demonstrate a technical procedure and recommend a narrower question. They should distinguish such coaching from supplying an unacknowledged solution. The student records consequential assistance in plain language after each substantive interaction, and the mentor confirms or corrects the entry. A brief oral explanation can clarify whether the student understands a revision; it is formative evidence, not a disciplinary interrogation.

Adult status creates unequal opportunities to disagree. At project initiation the student and mentor therefore rehearse a disagreement protocol: state the contested interpretation, identify the evidence that would change it, agree on a bounded next

step and preserve the result. Feedback should specify the intellectual standard and make a credible offer of support. Yeager and colleagues' experiments make interpretation of critical feedback a relevant design concern, but their findings are not a license to apply an identical phrase or infer equal effects in every setting [R8]. Mentor training should use actual examples and student feedback rather than merely completing a checklist.

Contribution transparency should avoid both honorary student authorship and erasure of genuine student work. Brand and colleagues describe a contribution-oriented approach to scholarly credit [R12]. For this pathway, a role declaration records question development, investigation, analysis, software, visualization, interpretation and writing, with examples of work performed. It informs authorship discussions without mechanically determining eligibility. Funding, seniority, supervision or platform ownership alone does not entitle an adult to authorship of every learner's paper. Student authors must be able to explain and accept responsibility for their contributions, with age-appropriate support and a route to challenge the credit record.

AI assistance belongs in the same accountability architecture. Learners may use approved tools to brainstorm, translate, draft code or inspect an explanation when the assignment permits it. The record should identify consequential uses, source checks and human decisions. Private student information and restricted research material must not be entered into unapproved services. A model's fluency cannot establish a finding, and the model is not assigned authorship or responsibility. The final public claim remains accountable to named human contributors; the learning assessment separately records what tools and help were available.

5. An assessment architecture that resists performance theatre

The principal measurement problem is whether a learner can investigate beyond the polished artifact. The proposed assessment presents an unfamiliar but accessible problem, a small evidence packet and a deliberately contestable preliminary claim. Students first give an initial response, then receive additional information containing both useful evidence and a plausible distraction. They may inspect a short methodological explanation before submitting a revised response and a change rationale. The task tests learning from new evidence and revising a claim under bounded conditions, rather than reproducing the project rehearsed with a mentor.

The proposed primary rubric has five dimensions scored from zero to four: question and construct clarity; appropriateness of evidence and comparison; treatment of uncertainty and rival explanations; substantive revision in response to information; and a warranted, audience-appropriate conclusion. The total ranges from zero to twenty. Each dimension requires behaviorally anchored examples. A high revision

score requires a justified response to evidence, including retaining an initial conclusion when the new information does not warrant change. Verbosity, rhetorical confidence, graphic polish and citation quantity receive no independent credit. Rubric development and pilot calibration must occur before confirmatory recruitment.

Two assessors, masked to allocation and stripped of school, author, mentor and venue identities, independently rate each response. An independent calibration set precedes operational rating; disagreement triggers a documented adjudication procedure without allowing assessors to infer allocation from project materials. Reliability is estimated with uncertainty for total scores and individual dimensions. If agreement is weak, the measurement problem is reported and the instrument revised for a later study; an unreliable confirmatory result is not rescued by choosing the most favorable rubric dimension after analysis.

Accessible administration is part of construct validity. Students may use approved dictation, screen readers or equivalent oral responses when these preserve the reasoning demands and are documented consistently across arms. All students receive the same task resource and, where needed, the same translation support. Parallel forms should be piloted for age suitability and comparable challenge. The primary task permits a common fixed source packet and approved accessibility tools, but no live generative assistant or individual mentor. This restriction concerns the defined independent-transfer estimand; it does not prohibit supported inquiry in the programme. Authentic project quality and AI-supported performance are assessed separately.

The separation of learning from rank is also consistent with Butler's experimental comparison of task-involving and ego-involving evaluation [R10]. The present design does not assume identical motivational effects in adolescent research or abolish grades by inference. It uses that distinction to specify a testable assessment policy: project feedback addresses evidence and decisions, whereas public status and venue labels are withheld from outcome assessors. Whether this arrangement improves student experience remains an empirical question.

6. Prospective trial: population, allocation and contrast

The proposed trial recruits students in grades eight through eleven through participating schools rather than selecting only students who already possess a research portfolio. Schools should publicize the pathway through ordinary enrollment channels, provide equipment and scheduling support, and record the eligible, approached, enrolled and consenting populations. Research participation is voluntary and distinct from receiving the educational programme. No recruitment, allocation or data collection has occurred for this protocol. Institutional review, safeguarding arrangements, registered procedures and a feasible budget are prerequisites to implementation.

A planning configuration comprises forty-eight intact inquiry cohorts across twelve schools, with approximately twenty students per cohort. These numbers are a logistical scenario, not a completed power calculation or a justified final sample. Rosters and baseline measures would be fixed before allocation. Within school and subject blocks, an independent statistician would randomize cohorts equally to structured Designership or an attention- and resource-matched inquiry pathway. Mentor eligibility, experience and availability are recorded before allocation. Dedicated mentor teams are assigned to cohorts from a common eligible pool through a prespecified procedure within school/subject blocks before the arm draw, preventing selective assignment of preferred mentors to Designership. Teams receive common safeguarding preparation and subsequently equal-duration preparation for their allocated pathway; experience and relevant expertise are reported by arm. Each team serves only one trial arm. If shared facilitators or scheduling make that separation impossible, the allocation unit must expand before registration and the power analysis must be repeated. Allocation concealment applies until enrollment and baseline assessment are complete.

Both pathways receive sixteen weeks of protected time: ninety scheduled minutes weekly, including eight thirty-minute mentor meetings across the programme, identical materials allowances and access to the same technical and safety resources. Both undertake an achievable inquiry project and receive competent general feedback. The Designership arm additionally follows explicit question-and-decision charters, a structured revision ledger, contribution verification, reciprocal peer critique and an editorial response exercise. The comparison arm uses the school's ordinary project organization within the matched provision. It retains essential ethical safeguards and student support; the contrast does not withhold safety, consent protection or basic research integrity.

The intervention therefore estimates the incremental value of a structured inquiry-and-accountability package over a supported inquiry opportunity. It does not isolate the causal effect of each component or test whether mentoring is preferable to no mentoring. Both arms may seek an appropriate public outlet after the primary assessment under the same editorial rules. Publication acceptance is never determined by allocation. Protected time, mentor contact, material use, intervention deviations and cross-arm exposure are recorded so that the intended contrast can be assessed. A badly implemented contrast cannot establish that the underlying educational theory has failed, but it can establish that the tested delivery strategy was ineffective.

7. Estimands, hypotheses and analysis

The primary estimand is the intention-to-treat mean difference in the sixteen-week unfamiliar-task total score between assignment to structured Designership and assignment to the matched pathway, averaged over the consenting, baseline-enrolled student population in participating cohorts. Assignment, rather than project completion or publication, defines inclusion. The primary hypothesis is a positive difference; the confirmatory test is two-sided at a prespecified five-percent level with a ninety-five-percent confidence interval. A null or negative estimate is a legitimate result. Generalization to nonparticipating schools requires additional assumptions and evidence.

The primary model is an individual-level analysis of covariance with baseline transfer score, grade and blocking variables specified before allocation, accounting for the cohort assignment unit. Adjusted arm means are standardized over the pooled baseline-enrolled student covariate distribution with equal student weights, so larger cohorts contribute proportionately to the student-average estimand rather than receiving the equal weight of a small cohort. Follow-up attendance and project completion do not determine weights. An equally weighted cohort-average estimate is a separately labeled sensitivity analysis; enrollment-size associations and any informative cluster-size concerns are reported. The analysis plan should choose a defensible small-sample correction and state its degrees of freedom. A block-respecting randomization analysis provides a robustness check when the realized number of cohorts permits it. The report will present the adjusted raw-score difference and a standardized difference using a prespecified baseline scale, with uncertainty for both. Students contribute through their allocated cohort even if their mentor or course placement later changes.

Secondary outcomes are independently rated project reasoning quality, six-month transfer on a parallel form, student-reported belonging and agency, documented ability to explain contribution, and time burden. Publication, submission and public uptake are descriptive dissemination measures. Equity analyses examine prespecified interactions with baseline opportunity indicators, including prior research access, economic disadvantage, disability-related accommodations and language support. They will report estimates and uncertainty rather than interpreting a nonsignificant interaction as proof of equal benefit. Small cells will be suppressed in public reporting. Unless adequately powered and registered, subgroup and mediation analyses remain exploratory.

Mechanism analyses may ask whether consequential revisions and student ownership explain outcome differences, but post-allocation mediator associations are not randomized mechanisms. They require explicit assumptions about mediator-outcome confounding and cannot establish causality simply by fitting a path diagram. Per-protocol analyses based on fidelity are secondary and vulnerable to selection. The primary result should therefore remain visible even if an engaged subgroup performs

well. CONSORT's cluster-trial extension provides relevant requirements for describing cluster allocation, recruitment, flow and analysis; following a reporting checklist does not by itself repair design weaknesses [R11].

8. Precision, missingness and falsification

Before confirmatory recruitment, investigators should conduct an external pilot focused on feasibility and measurement, using participants who will not enter the confirmatory sample. The pilot estimates task reliability, plausible score dispersion, cohort correlation, recruitment variation and attrition without selecting a favorable intervention effect. A simulation-based power analysis should vary the number and sizes of cohorts, intraclass correlations, baseline prognostic strength, unequal enrollment and realistic missingness. The final sample is chosen to estimate a stakeholder-defined minimum educationally meaningful difference with adequate precision, not to manufacture significance for an arbitrary expected effect. Simulation code and assumptions should be public before outcomes are observed.

The enrollment roster supplies a denominator for follow-up. Investigators record reasons for missing assessments separately from reasons for leaving the programme, offer accessible assessment appointments and continue permitted follow-up after course withdrawal. The main missing-data strategy should use a prespecified multilevel model or compatible multiple imputation incorporating allocation, baseline performance, school or cohort structure and predictors of missingness. Complete-case estimates are reported as sensitivity analyses. Because missing-at-random is untestable from observed data alone, pattern-mixture or tipping-point analyses should show how conclusions change under plausible poorer unobserved outcomes, including arm-specific departures. Missing public portfolios are not automatically scored as missing learning outcomes.

The framework faces three substantive falsification tests. First, better project presentation without improvement on the unfamiliar-task outcome supports an artifact-production explanation rather than the proposed independent-transfer benefit. Second, higher apparent contribution accompanied by adult takeover in explanation checks undermines the ownership claim. Third, an average benefit accompanied by increased unpaid burden, exclusion or privacy harm requires redesign even when the primary estimate is positive. These tests must be interpreted with measurement uncertainty; they are not invitations to declare failure from every nonsignificant secondary result.

Progression decisions should be agreed with students, families, teachers and an independent oversight group before enrollment. They should include minimum measurement reliability, feasible delivery, unacceptable safeguarding events and a cap on additional out-of-class work. The thresholds need pilot evidence and local deliberation; none is presented as already agreed here. A staged decision can

recommend revision or a further adequately powered trial instead of immediate scale. Public registration should preserve the original plan, dated amendments, analysis code and a plain-language explanation of deviations.

9. Child privacy, consent and editorial independence

Education, research participation and public dissemination require distinct decisions. Families should be able to accept instruction without consenting to research use of records, and a student should be able to participate in the evaluation without creating a permanent public identity. Where applicable, an authorized ethics body determines the governing requirements and any permissible waivers before research begins. United States HHS rules distinguish child assent from parental permission and provide additional protections for covered research involving children [R13]. The protocol adopts meaningful student assent and appropriate guardian permission as design commitments without claiming that this manuscript constitutes legal or institutional approval.

The default research dataset uses pseudonymous identifiers, separate identity keys, role-limited access and a documented retention and deletion schedule. Public examples should use synthetic or independently consented material; removing names alone may not prevent recognition in a small school. Location, health, family circumstances, voices, faces and third-party observations require particular attention. Research datasets cannot become automatic retrieval material for public paper agents. Only a deliberately approved public artifact and its lawful sources may enter that reader-facing collection, and passage-level provenance should distinguish student claims from source claims.

Public authorship and an ORCID record are optional dissemination choices, not conditions for receiving educational credit. A minor who declines a searchable public profile should have an equivalent private completion route. Consent for a particular publication is requested after the artifact and its identifying content can be reviewed. Families and students need a plain explanation that downloaded material and external identifiers may persist even if a local page is removed. Private review, delayed release and appropriate de-identification are viable outcomes; an adult's enthusiasm for discoverability does not override the learner's interests.

When programme leaders own the publication platform or author scholarship evaluating it, those relationships must be disclosed. Reviewers and handling editors for that evaluation should be independent of the intervention's commercial or institutional success. AI-assisted critique can help identify weak arguments, but it does not constitute independent human peer review. The present manuscript contains no assertion that a trial has been ethically approved, registered, conducted or externally peer reviewed. Human editors must separately determine the publication category, review record and author responsibilities for the final edition.

10. Equity, resources and costs

Designership can reproduce advantage if access depends on family introductions to scientists, expensive devices, discretionary time or confidence in academic English. The pathway therefore requires an opportunity audit before recruitment: who receives an invitation, who can attend, which topics need specialist facilities, whose language practices are recognized and who must relinquish paid or caring work to participate? The audit concerns conditions schools can change. It should not turn demographic information into a ranking of presumed student potential.

The minimum viable model allocates meaningful work within the timetable, supplies necessary materials, offers accessible formats and compensates staff rather than relying invisibly on volunteer labor. Projects should be designed so that a learner with limited home broadband can complete core inquiry at school. Community knowledge may identify consequential questions without being extracted as free data. Where a project involves community partners, the agreement specifies the purpose, permissible uses, feedback route and whether the partner wishes to be named. Refusing publicity is not evidence of low engagement.

Costs should be reported from both school and family perspectives: mentor preparation, contact time, coordination, assessment, software, equipment, accessibility and students' out-of-class time. A cost-consequence analysis can place these alongside learning and burden outcomes without prematurely pricing a publication or assigning monetary value to a child's profile. If a structured pathway consumes more resources than its comparison, the matched contrast and realized spending must be reported. Institutional prestige and social-media impressions should never obscure who bore the work or whether an equally effective lower-cost arrangement was available.

11. Implications for schools and scholarly publishing

The proposed framework makes a journal a place for disciplined public contribution while keeping schooling responsible for learning that publication cannot certify. Its signature is an inspectable relationship among question, evidence, revision, contribution and judgment. A careful school-based inquiry can enter that relationship without pretending to be a definitive scientific discovery. An ambitious finding can be rejected or narrowed without treating the learner as unsuccessful. Editorial rigor and developmental encouragement are compatible when they concern the work rather than the status of the author.

For The Research University and Designership Press, the immediate infrastructure implication is separation of learning assessment, confidential editorial workflow and public publication metadata. The reader should see what was investigated, who did what, which sources support the claim and which version is being read. The student should see clear decisions, constructive revision requests and an accessible

explanation of publication choices. The institution should be able to audit consent, conflicts, source integrity and correction history without exposing private educational records. A paper agent can assist navigation through approved evidence; responsibility for interpretation and publication remains human.

The distinctive research question is consequently modest and demanding: does structured, mentored inquiry with accountable revision improve transferable reasoning beyond an equally supported inquiry opportunity, at acceptable cost and without narrowing access? The prospective design specifies how that question could be answered and what would count against the proposal. It offers no invented participants, effect sizes or approval record. Designership becomes a credible educational contribution when its claims survive such tests, its failures remain visible and its learners retain both intellectual responsibility and the freedom to choose how their work becomes public.

References and verification record

References retain their original authorship and publication identity. Verification dates below record source and metadata checking; they do not confer review or approval on this article.

[R1] Brigid J.S. Barron; Daniel L. Schwartz; Nancy J. Vye; Allison Moore; Anthony Petrosino; Linda Zech; John D. Bransford; The Cognition and Technology Group at Vanderbilt (1998). Doing With Understanding: Lessons From Research on Problem- and Project-Based Learning. Journal of the Learning Sciences.

DOI: https://doi.org/10.1080/10508406.1998.9672056

Source: https://doi.org/10.1080/10508406.1998.9672056

Verified 2026-10-16. Crossref title and individual author order checked; institutional author PDF and Stanford author bibliography additionally identify The Cognition and Technology Group at Vanderbilt, retained here although omitted from Crossref author list.

[R2] Joseph A. Taylor; Catrina T. Adams; Anne L. Westbrook; Jodi Creasap Gee; Jessaca K. Spybrook; Susan M. Kowalski; April L. Gardner; Mark Bloom (2022). The effect of a student-teacher-scientist partnership program on high school students' science achievement and attitudes about scientists. Journal of Research in Science Teaching.

DOI: https://doi.org/10.1002/tea.21733

Source: https://doi.org/10.1002/tea.21733

Verified 2026-10-16. Title, author order and publication year checked against Crossref registration metadata; claims restricted to original study or conceptual/methodological scope. Taylor et al. verified against publisher page; online-first 2021, issue 2022.

[R3] Louise Puslednik; Patrick C Brennan (2020). An Australian-based authentic science research programme transforms the 21st century learning of rural high school students. Australian Journal of Education.

DOI: https://doi.org/10.1177/0004944120919890

Source: https://doi.org/10.1177/0004944120919890

Verified 2026-10-16. Title, author order and publication year checked against Crossref registration metadata and publisher/institutional source where accessible.

[R4] Mercedes Edry; Irit Sasson; Yehudit Judy Dori (2023). Secondary School Apprenticeship Research Experience: Scientific Dispositions and Mentor-Student Interaction. Education Sciences.

DOI: https://doi.org/10.3390/educsci13050441

Source: https://doi.org/10.3390/educsci13050441

Verified 2026-10-16. Title, author order and publication year checked against Crossref registration metadata and publisher/institutional source where accessible.

[R5] Daniel L. Schwartz; Taylor Martin (2004). Inventing to Prepare for Future Learning: The Hidden Efficiency of Encouraging Original Student Production in Statistics Instruction. Cognition and Instruction.

DOI: https://doi.org/10.1207/s1532690xci2202_1

Source: https://doi.org/10.1207/s1532690xci2202_1

Verified 2026-10-16. Title, author order and publication year checked against Crossref registration metadata and publisher/institutional source where accessible.

[R6] Manu Kapur (2008). Productive Failure. Cognition and Instruction.

DOI: https://doi.org/10.1080/07370000802212669

Source: https://doi.org/10.1080/07370000802212669

Verified 2026-10-16. Title, author order and publication year checked against Crossref registration metadata and publisher/institutional source where accessible.

[R7] Kwangsu Cho; Charles MacArthur (2010). Student revision with peer and expert reviewing. Learning and Instruction.

DOI: https://doi.org/10.1016/j.learninstruc.2009.08.006

Source: https://doi.org/10.1016/j.learninstruc.2009.08.006

Verified 2026-10-16. Title, author order and publication year checked against Crossref registration metadata and publisher/institutional source where accessible.

[R8] David Scott Yeager; Valerie Purdie-Vaughns; Julio Garcia; Nancy Apfel; Patti Brzustoski; Allison Master; William T. Hessert; Matthew E. Williams; Geoffrey L. Cohen (2014). Breaking the cycle of mistrust: Wise interventions to provide critical feedback across the racial divide.. Journal of Experimental Psychology: General.

DOI: https://doi.org/10.1037/a0033906

Source: https://doi.org/10.1037/a0033906

Verified 2026-10-16. Title, author order and publication year checked against Crossref registration metadata and publisher/institutional source where accessible.

[R9] Carla Herrera; Jean Baldwin Grossman; Tina J Kauh; Jennifer McMaken (2011). Mentoring in Schools: An Impact Study of Big Brothers Big Sisters School-Based Mentoring. Child Development.

DOI: https://doi.org/10.1111/j.1467-8624.2010.01559.x

Source: https://doi.org/10.1111/j.1467-8624.2010.01559.x

Verified 2026-10-16. Title, author order and publication year checked against Crossref registration metadata and publisher/institutional source where accessible.

[R10] Ruth Butler (1988). Enhancing and undermining intrinsic motivation: The effects of task-involving and ego-involving evaluation on interest and performance. British Journal of Educational Psychology.

DOI: https://doi.org/10.1111/j.2044-8279.1988.tb00874.x

Source: https://doi.org/10.1111/j.2044-8279.1988.tb00874.x

Verified 2026-10-16. Title, author order and publication year checked against Crossref registration metadata and publisher/institutional source where accessible.

[R11] Marion K. Campbell; Gilda Piaggio; Diana R. Elbourne; Douglas G. Altman; CONSORT Group (2012). Consort 2010 statement: extension to cluster randomised trials. BMJ.

DOI: https://doi.org/10.1136/bmj.e5661

Source: https://doi.org/10.1136/bmj.e5661

Verified 2026-10-16. Title, author order and publication year checked against Crossref registration metadata and publisher/institutional source where accessible.

[R12] Amy Brand; Liz Allen; Micah Altman; Marjorie Hlava; Jo Scott (2015). Beyond authorship: attribution, contribution, collaboration, and credit. Learned Publishing.

DOI: https://doi.org/10.1087/20150211

Source: https://doi.org/10.1087/20150211

Verified 2026-10-16. Title, author order and publication year checked against Crossref registration metadata and publisher/institutional source where accessible.

[R13] U.S. Department of Health and Human Services, Office for Human Research Protections (n.d.). 45 CFR 46, Subpart D-Additional Protections for Children Involved as Subjects in Research. HHS Office for Human Research Protections.

Source: https://www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/common-rule-subpart-d/index.html

Verified 2026-10-16. Official current regulatory text, especially definitions in §46.402 and assent/parental-permission requirements in §46.408; no publication date asserted, accessed 16 October 2026.

[R14] Clark A. Chinn; Betina A. Malhotra (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. Science Education.

DOI: https://doi.org/10.1002/sce.10001

Source: https://doi.org/10.1002/sce.10001

Verified 2026-10-16. Title, author order and publication year checked against Crossref registration metadata and publisher/institutional source where accessible.

Source and authorship notes

The uploaded seed document is a commissioning brief, not independent evidence for historical, program-outcome or reputation claims. Movement Thinking, SIDE and Design Realization Garages are identified as proposed author frameworks.

This is original conceptual and protocol writing, not a republication or a systematic review. No trial data, invented results, named reviewers, ethics number, registration number or DOI is supplied.

Metadata verified 16 October 2026. Taylor et al. appeared online 10 November 2021 in a 2022 issue; displayed bibliography uses issue year. HHS regulatory text has no asserted publication year; accessed 16 October 2026.

Source audit includes raw Crossref metadata; several publisher pages block automated full text, so argument uses only accessible abstracts/primary institutional versions and does not claim comprehensive appraisal of inaccessible full texts.

Human authorship and editorial responsibility require actual author review. AI-assisted drafting/research verification is not independent human peer review.

D04 / Equitable infrastructure

A School Must Be Reachable: Digital Infrastructure, Accessible Opportunity, and the Designership Service Obligation

Dr. Freedom Cheteni

The Research University | Volume 1, 2026 | Edition 1.0 | D04

Evidence-informed policy analysis and prospective comparative implementation protocol

Original analyses and prospective protocols. New independent human peer review is not recorded for this edition.

DOI: not assigned to this research edition.

Read online | Author ORCID: 0009-0006-3507-0094 | Editorial record

Abstract

Digital and spatial schools are often evaluated through platform availability, visual sophistication, or participation counts. These measures do not establish that learners can enter consequential intellectual work, obtain human support, demonstrate understanding, and continue after disruption. This evidence-informed policy analysis develops an original infrastructure-to-opportunity chain connecting availability, usable entry, supported participation, intellectual work, credible demonstration, and continuity. It proposes a minimum viable learning service and meaningful exit protections, treating offline alternatives, accessibility, household burden, and public-publication choice as design responsibilities. A prospective six-district, twenty-four-school protocol evaluates a supported Designership inquiry package above a common essential-service floor through a district-stratified matched-pair randomized delayed-start design. Its confirmatory hierarchy distinguishes blinded science-transfer performance, participation, and access success; its economic component reports district and household cost-consequences. Randomization, aligned assessment timing, interference, missingness, subgroup uncertainty, and ethical protections are specified, with an assumption-dependent matched-comparison fallback chosen before outcome inspection. No new intervention data or effectiveness results are reported. The contribution is a falsifiable service framework for governing educational infrastructure by the opportunities it sustains rather than the technologies it displays.

Keywords: Designership, K-12 digital infrastructure, educational equity, accessibility, learning service, matched-pair cluster protocol, cost-consequence analysis, offline learning

1. From a digital campus to an educational service

The consequential question for a digital school is not whether its environment resembles an impressive campus. It is whether a learner can enter a worthwhile task, obtain appropriate support, demonstrate understanding, and continue learning when a device, connection, platform, or personal circumstance fails. A spatial interface may enable an otherwise inaccessible investigation; an ordinary telephone may provide the more effective route to another objective. Designership, as developed here, makes the organization accountable for this relationship between educational purpose and the conditions of participation. Its object is a dependable learning service rather than a predetermined technological form.

The distinction matters for K-12 systems considering virtual campuses, remote research mentorship, and publication of student work. A functioning website is an institution-facing accomplishment. An accessible opportunity to learn is a student-facing accomplishment. These can diverge: a technically available class may depend on uninterrupted broadband, an adult at home, expensive headphones, rapid keyboard entry, or a willingness to display a bedroom. None of these conditions should disappear from an equity assessment merely because they sit outside the school's procurement contract. The infrastructure boundary must follow the learning task into the circumstances in which it is undertaken.

This article develops an original policy framework and a prospective comparative evaluation protocol. It reports no implementation trial, recruited participants, new outcome data, or demonstrated intervention effect. Its evidence base is a purposive selection of official datasets, technical standards, and primary studies, checked against publisher or institutional records; it is not a systematic review. “Educational service obligation” denotes the stewardship standard proposed here, not a new interpretation of statutory duties. Existing legal requirements require separate, jurisdiction-specific review. The contribution is an inspectable infrastructure-to-opportunity chain, a minimum service floor, and an evaluation that can reject attractive but ineffective designs.

2. What the evidence permits-and what it does not

UNESCO's 2023 Global Education Monitoring Report asks whether educational technology serves educational purposes under appropriate conditions, rather than assuming that deployment demonstrates benefit [R1]. U.S. household-access statistics similarly distinguish computer access from access only through a smartphone [R2]. A district should therefore avoid a binary “connected” indicator that erases the difference between reading a message and manipulating a dataset. These sources motivate attention to access conditions; they do not estimate the effect of the service package proposed in this article.

Primary experiments caution against treating hardware as an instructional intervention. Fairlie and Robinson's randomized provision of home computers increased ownership and use without improving the educational outcomes assessed in their California study [R3]. Cristia and colleagues' Peruvian school trial likewise found greater computer access and use without evidence of improved mathematics or language scores over its reported follow-up [R4]. These results do not establish that computers never help. They establish that removing one input constraint need not repair the instructional pathway between access and learning.

Other experiments show why the pathway requires specification. Muralidharan, Singh, and Ganimian evaluated a particular personalized after-school instructional program in urban India and found learning gains [R5]. Angrist, Bergman, and Matsheng found benefits from a combined telephone-and-SMS intervention supporting children through parents during school closure in Botswana [R6]. Neither is evidence that a U.S. virtual school, an immersive campus, or an automated research agent will reproduce those effects. Together, the studies justify comparing complete instructional arrangements, including human support and task design, instead of ranking technologies by apparent sophistication.

Immersion also deserves a testable claim. Makransky, Terkildsen, and Mayer's science-laboratory experiment found greater reported presence alongside poorer learning in its immersive condition [R7]. Its setting and learners limit generalization to K-12 classrooms. The relevant inference is methodological: presence, satisfaction, and visual realism cannot substitute for learning measures. A spatial environment should identify the representation, interaction, or collaboration that requires three dimensions, state the expected benefit, and retain an accessible alternative capable of supporting the educational objective.

3. The infrastructure-to-opportunity chain

The proposed chain contains six distinct links: availability, usable entry, supported participation, consequential intellectual work, credible demonstration, and continuity. Availability asks whether the necessary resources exist at the relevant place and time. Usable entry asks whether the learner can actually authenticate, perceive instructions, navigate, and begin. Supported participation asks whether feedback and assistance arrive before a difficulty becomes exclusion. Intellectual work asks whether the task requires explanation, investigation, design, or justified revision. Demonstration asks whether the resulting evidence supports an interpretation of learning. Continuity asks whether that opportunity survives an interruption or institutional transition.

This is a diagnostic causal hypothesis, not a validated scale. A district might provide devices while authentication remains inaccessible; improve login success while assignments remain intellectually thin; or support excellent work while students lose

all evidence when a subscription ends. Each failure occurs at a different link and calls for a different response. The chain prevents a single utilization dashboard from collapsing these mechanisms into “engagement.” It also locates Movement Thinking's concern with consequential questions and preserved revision within the practical conditions that allow students to act on them.

The chain predicts that infrastructure investments will have heterogeneous educational value because the binding constraint differs across learners and tasks. Adding bandwidth may matter greatly for a student using a shared telephone and little for one whose difficulty is receiving comprehensible feedback. Conversely, a strong mentor cannot remedy an inaccessible assessment interface by encouragement alone. The policy implication is to identify barriers before prescribing solutions, then test whether removing the identified barrier changes participation and learning. The proposal does not require every link to be digital or to be supplied by one vendor.

4. A minimum viable learning service and meaningful exit

The minimum viable learning service is a proposed floor beneath innovation. Before making a digital or spatial task compulsory, a school should document an appropriate device or equivalent route; a workable connection or offline delivery arrangement; accessible task instructions and assessment; a named human contact; a feasible feedback schedule; and a continuity plan. “Equivalent” concerns access to the learning objective and fair assessment, not identical screen experiences. The floor is deliberately broader than software uptime. A system that operates continuously but cannot supply a usable task to a particular learner has not delivered the promised service.

Every core task should have a route map stating what can be downloaded, completed without continuous connectivity, submitted later, or undertaken with school-provided support. Offline work must include a return channel for feedback; a packet left at home is not a complete service. Alternative routes should carry comparable intellectual expectations and timely teacher attention. Requiring a family to negotiate an exception repeatedly turns access into a test of advocacy. Schools should make ordinary routes discoverable in advance, while preserving individualized accommodations and avoiding disclosure of a student's disability to peers.

Meaningful exit includes changing interface, declining optional immersion, leaving the research component, or transferring institutions without forfeiting earned credit or access to one's permitted work. Procurement should specify exportable learner artifacts, readable feedback, version history where appropriate, and a documented transition procedure. These are proposed contractual protections, not claims that all current systems already provide them. Exports must respect other students' privacy and source licenses. Public publication must remain separate from course completion:

a student should be able to demonstrate scholarship privately without accepting searchable, persistent exposure as the price of an education.

5. Accessibility is a task property as well as an interface property

WCAG 2.2 provides testable web-accessibility criteria and explicitly acknowledges that conformance does not address every user need [R8]. It should inform procurement, authoring, and testing, but an automated scan cannot certify that a lesson is usable. A prospective service audit should follow complete tasks, including authentication, finding instructions, handling an error, uploading work, receiving feedback, and exporting an artifact. Human evaluation with relevant assistive technologies is indispensable. Findings should identify the task, configuration, severity, remedy, and retest date rather than produce an unqualified accessibility badge.

For immersive environments, the W3C XR Accessibility User Requirements document identifies issues such as multimodal interaction, motion-agnostic controls, orientation, and alternatives to spatial audio [R9]. It is a Working Group Note, explicitly not a baseline conformance standard or an endorsement of every proposal. Its role here is to structure questions for design and user testing. A headset experience should have an immediate exit, a non-headset route, and a way to communicate discomfort privately. Accessibility should be evaluated before assignment, not discovered when an individual student fails to participate.

CAST's Universal Design for Learning Guidelines 3.0 offer design considerations concerning engagement, representation, and action and expression [R15]. They support planning for learner variability; citing them does not validate a particular curriculum. The proposed protocol therefore separates three questions: whether the interface meets specified criteria, whether learners can complete the task through their chosen route, and whether they learn. Schools should preserve the distinction even when a single purchasing decision addresses all three. A beautiful interface can satisfy neither task accessibility nor the intended educational purpose.

6. Human support, open resources, and institutional accountability

The service floor requires staffing commitments. A named teacher remains responsible for interpreting learning evidence; a support contact helps resolve access failures; and an administrator owns unresolved barriers. Response schedules should distinguish urgent exclusion from ordinary instructional questions and should fit paid working time. An automated assistant may locate a source passage or help formulate a question, but it must provide a clear route to a person when an answer affects assessment, safeguarding, accommodation, or participation. Availability of an assistant does not demonstrate availability of accountable educational support.

Open educational resources can improve adaptability and reduce some licensing constraints. UNESCO's 2019 Recommendation defines their openness through permissions attached to public-domain or openly licensed materials [R10]. A repository's public visibility does not establish permission to translate, adapt, redistribute, or build a paper agent around every item. The proposed service maintains a rights register at resource and version level, preserves original authorship, and distinguishes a new student analysis from a republication. Course publication should add accountable interpretation and inspectable evidence rather than manufacture originality through cosmetic rewriting.

The 2024 U.S. National Educational Technology Plan distinguishes access, design, and use divides [R16]. This framework is compatible with the proposed chain, while the chain adds explicit demonstration and continuity tests. A school board's public service statement should identify the learning purposes, support responsibilities, available alternatives, evidence reviewed, and unresolved barriers. It should not report adoption totals as educational success. Institutional accountability becomes more credible when families can see how an access problem is resolved and how an ineffective investment can be changed or discontinued.

7. Full lifecycle costs and the distribution of burden

The appropriate economic unit is a supported opportunity to learn over a specified period, not the advertised price of a license. Costs include devices, replacement cycles, repair, connectivity, storage, cybersecurity, accessibility work, content preparation, onboarding, help-desk labor, teacher planning, mentorship, assessment, and transition. Spatial provision adds cleaning, fitting, supervision, room preparation, and alternatives for learners who cannot use headsets. Reusing equipment can reduce procurement costs while increasing maintenance. The analysis should therefore state the service configuration and expected useful life instead of assuming a universal price per pupil.

A district perspective alone can conceal transfers to households. Families may supply data plans, electricity, transport to reliable access, translation, supervision, quiet space, and unpaid troubleshooting. Teachers may absorb unrecorded preparation or support time. The proposed economic evaluation reports district expenditure and a wider school-and-household perspective separately. Time is recorded in natural units before assigning any monetary value; alternative valuations are shown because an hour of family time should not become less important merely because household earnings are low. Costs borne by excluded learners must remain visible.

The primary presentation is cost-consequence: incremental resources alongside learning, participation, accessibility, burden, and adverse-event outcomes. It does not compress dignity or access into a speculative financial return. CHEERS 2022 informs transparent reporting of perspective, time horizon, quantities, valuation, assumptions, and uncertainty, while originating in health economics rather than validating this educational design [R14]. Capital costs should be annualized under explicit useful-life assumptions. Scenarios should vary device life, staffing intensity, enrollment, replacement prices, and real discount rates of 0%, 3%, and 5%, with local policy rates also reported.

8. Prospective study: distinguish the service floor from the enhancement

The proposed study recruits six districts and four secondary schools per district, subject to feasibility, agreements, and an independent ethics determination before any recruitment. Grades 8-9 science provide a bounded curricular context. These are planning targets, not an achieved sample. Before baseline, all participating schools remedy urgent access and accessibility failures and adopt the minimum service floor. Essential support is never randomized away. The randomized contrast estimates the additional effect of a supported Designership learning package above this common floor; it cannot estimate the effect of establishing the floor itself.

The enhancement comprises one twelve-week inquiry module, protected teacher co-design time, a named research mentor, accessible low-bandwidth materials, and structured feedback on successive student explanations. Spatial activities are optional representations within the module, with non-headset equivalents; the trial does not isolate a headset effect. Within each district, its four schools form two matched pairs using baseline attainment, enrollment, access constraints, and available support. An independent statistician randomizes one school per pair to immediate enhancement and the other to delayed start. All twelve pairs run concurrently over the same school term; allocation follows baseline enrollment and assessment.

The control schools retain their ordinary curriculum and the common service floor during the comparison term. Both groups complete the same baseline and end-of-term transfer assessments in aligned calendar weeks. Immediate-start schools deliver the

module once before the primary endpoint. Delayed schools may begin it only after that endpoint. A later follow-up may describe durability and implementation in both groups, but cannot preserve the original untreated contrast. This matched-pair parallel design avoids the repeated exposure and time-since-crossover ambiguities of a multi-term rollout. Stepped-wedge designs can be useful, but require explicit treatment of time and correlation [R12].

Schools may share teachers, curriculum, or support, creating interference across assigned groups. Cross-school staffing and material sharing will be mapped before assignment and recorded during the study. Necessary assistance is never restricted to protect a contrast. If district-wide delivery makes school-level separation implausible, the study must instead recruit enough districts for a district-cluster comparison or remain a feasibility study. Its design must change before registration. An impressive pupil count cannot substitute for independent assignment units, and contamination cannot be solved by silently reclassifying schools according to observed compliance.

9. Outcomes, recruitment, and precision

The confirmatory hierarchy begins with learning: blinded scoring of a common science-transfer task at the end of a twelve-week module, adjusted for a parallel baseline task. The rubric assesses use of evidence, explanation of a mechanism, consideration of an alternative, and justified revision. Pilot work outside the analytic sample will examine task accessibility, curricular fit, inter-rater agreement, and score behavior across routes; the rubric and scoring rules are then frozen. Familiarity with a particular interface must not become the construct being assessed. Scorers receive de-identified artifacts and remain unaware of allocation and school.

The second confirmatory outcome is participation: the proportion of the baseline research cohort completing at least four of five common curricular milestones offered in both arms with the same learning goals, timing, and minimum evidence requirements. The milestone denominator is frozen before allocation; unrelated platform events and additional enhancement activities do not enlarge it. The third is access success: the proportion of that cohort completing the same standardized entry-to-feedback task, under each assigned service arrangement, through the learner's preferred supported route without an unresolved blocking barrier. These thresholds are proposed operational choices, not validated natural cutoffs. Accessibility also remains an ethical monitoring constraint irrespective of statistical testing. Screen time, logins, presence, and satisfaction are descriptive rather than learning proxies. Family burden, teacher workload, and discomfort are prespecified secondary outcomes.

Eligibility includes pupils on prespecified target-class rosters. Invitations, permission, assent, and baseline measurement precede allocation, creating a closed consenting baseline cohort for the primary analysis. Approximately sixty eligible learners per

school is an operational planning figure, not a power calculation or enrollment claim. Research participation requires guardian permission and age-appropriate assent, unless an independent ethics body approves a legally permissible alternative. Nonparticipants retain ordinary learning, assessment alternatives, and support. The same cohort is invited to both research assessments; transfers and withdrawals are documented without replacing missing pupils with new entrants. New entrants may join instruction and a separately described implementation cohort, but not the primary analytic cohort.

Simulation-based precision and power analysis will precede recruitment. It will vary school intraclass correlation, matched-pair similarity, baseline-outcome correlation, unequal school sizes, nonresponse, and contamination using independently justified ranges. Twelve matched pairs permit a finite randomization distribution, but cannot guarantee detection of a small educational effect. The protocol will publish its minimum detectable effect and assumptions. If feasible recruitment yields inadequate precision, the study will be described as a pilot and its uncertainty retained. Recruitment totals and participant flow will be distinguished from the number of genuinely independent clusters.

10. Analysis, missingness, and competing explanations

The primary estimand is the intention-to-treat effect of immediate enhancement assignment on end-of-term learning in the baseline cohort. A prespecified baseline-adjusted model includes matched-pair effects and school-level correlation; pair-respecting randomization inference accompanies model-based confidence intervals. The primary contrast gives equal weight to matched pairs, estimating the average pair-level contrast across participating school settings. A prespecified pupil-weighted analysis addresses the different policy question of average benefit across enrolled learners. With only twelve pairs, uncertainty remains material. Allocation, analysis rules, and treatment of unequal school size are fixed before outcome access.

Only the common comparison term supplies the confirmatory treatment contrast. Baseline and endpoint tasks use parallel forms, counterbalanced within schools and equally across assigned groups; no pupil repeats the same items. Timing is aligned across immediate and delayed schools. Implementation measures record actual mentor contact, teacher planning time, route use, and delivery delays; these describe the assigned package rather than replace assignment in the primary analysis. Exploratory compliance-adjusted or mediation analyses require additional assumptions. Attention and added adult time are components, so this trial cannot isolate a technology effect or validate an entire philosophy.

If randomization proves infeasible, a distinct matched-adoption study requires a revised timeline and separate prospective registration. It would collect at least three

comparable pre-adoption measurement periods before using group-time estimates with not-yet-treated schools and conditional parallel trends, informed by Callaway and Sant'Anna [R13]. Matching cannot remove unobserved leadership motivation. Overlap, anticipation, prior trends, concurrent reforms, and enrollment changes require examination; a nonsignificant pretrend does not prove identification. This fallback cannot be chosen after examining treatment outcomes or presented as equivalent to random assignment. Its claims remain explicitly assumption-dependent.

Missing assessments require explanation by route, school, baseline attainment, and reported barrier. An inaccessible assessment is an access failure and missing learning evidence, not proof of low ability. The analysis will report missingness, use prespecified multiple-imputation models when defensible, and show pattern-mixture and worst-plausible-case bounds. Confirmatory tests follow the stated hierarchy with multiplicity control specified at registration; estimates and intervals are reported regardless of significance. Sensitivity analyses will examine contamination and outcome availability without concealing the original intention-to-treat result. An intervention that drives pupils out of observable assessment cannot claim success only among those remaining.

11. Equity, privacy, and a publication boundary for minors

Prespecified heterogeneity examines baseline access constraints, disability-related support needs, socioeconomic disadvantage, language support, and rurality where samples permit safe interpretation. These are not interchangeable categories or biological explanations for performance. Analysis should estimate differences in effects with uncertainty and report small-cell suppression. An average gain does not establish equity if some learners face higher barriers or family costs. Small subgroups may be too imprecisely estimated to establish safety; qualitative access audits and individual remedies remain necessary. The absence of a statistically significant interaction is not evidence that all groups benefit equally.

The data plan uses the least information needed for the stated questions. Student identifiers remain in a separate district-controlled linkage file; research extracts exclude bedroom video, raw voice recordings, gaze traces, biometric inference, and unrelated behavioral telemetry. The U.S. Department of Education's guidance on online educational services provides an authoritative starting point for evaluating privacy arrangements, not a universal legal clearance [R11]. Applicable law, contractual controls, retention, access roles, breach procedures, and cross-border transfers require current local review before collection. Human support and ordinary grades cannot depend on research consent.

Public scholarship needs a separate boundary. A school may assess a student's argument, source use, and revision privately; public release requires a distinct, understandable choice with an age-appropriate process and an explanation of

persistence and searchability. ORCID participation, a public profile, or publication must not be a prerequisite for the compulsory learning activity. An article agent should retrieve only approved public material, identify its source passages, and avoid drawing on private learning records. Copyright, student authorship, safeguarding, and scientific review remain human responsibilities even when assistance makes dissemination easier.

12. Decision rules, limitations, and the contribution to Designership

A decision panel including students, family representatives, teachers, accessibility specialists, and district leaders will set affordability and educational-relevance thresholds before seeing results. Scaling requires an interpretable learning estimate, acceptable participation, no unresolved serious access barrier attributable to the design, and a credible funded support plan. A positive learning average cannot override a known preventable exclusion. A package may warrant revision if participation rises without learning, if gains depend on unsustainable unpaid labor, or if a simpler route delivers comparable outcomes with lower burden. A null estimate should prompt examination of implementation and precision without being relabeled success.

The main limitations are deliberate and consequential. The evidence selection is purposive; its studies come from different populations and cannot yield a pooled effectiveness claim. The service floor is a normative design contribution, not a validated checklist. The proposed trial bundles instructional and organizational changes and cannot isolate each mechanism. Secondary-school science does not represent early childhood, every subject, or every national system. The proposed school-level assignment remains vulnerable to spillover, nonresponse, and implementation problems; its small number of pairs limits precision. Cost projections and subgroup estimates will require cautious interpretation. None of these limits can be repaired by architectural elegance or institutional branding.

Designership earns practical meaning when ambition carries obligations that can be inspected. A digital school should make entry feasible, intellectual work worthwhile, support accountable, evidence credible, and exit possible. The proposed infrastructure-to-opportunity chain translates that position into observable service questions and a comparative research program. Its success would not be a claim that virtual space replaces school. It would be a demonstrated capacity to expand serious learning without transferring hidden costs and exclusions to the learners least able to absorb them. The next contribution is implementation under this protocol, with results reported whether they support or challenge the design.

References and verification record

References retain their original authorship and publication identity. Verification dates below record source and metadata checking; they do not confer review or approval on this article.

[R1] UNESCO Global Education Monitoring Report Team (2023). Global education monitoring report 2023: Technology in education: A tool on whose terms?. UNESCO.

Source: https://www.unesco.org/gem-report/en/publication/technology

Verified 2026-10-16. Official report landing page and report text; governance/access/evidence framing. Not a causal evaluation of the proposed service.

[R2] National Center for Education Statistics (2025). Percentage distribution of children ages 3 to 18, by whether they have home internet access, whether they have access through computer or only smartphone, and selected child and family characteristics: 2022 and 2023. Digest of Education Statistics, Table 702.12.

Source: https://nces.ed.gov/programs/digest/d25/tables/dt25_702.12.asp

Verified 2026-10-16. Official table; distinguishes computer access and smartphone-only access. Data years 2022-2023, table edition 2025. No numerical prevalence claim made.

[R3] Robert W. Fairlie; Jonathan Robinson (2013). Experimental Evidence on the Effects of Home Computers on Academic Achievement among Schoolchildren. American Economic Journal: Applied Economics, 5(3), 211-240.

DOI: https://doi.org/10.1257/app.5.3.211

Source: https://www.aeaweb.org/articles?id=10.1257/app.5.3.211

Verified 2026-10-16. Publisher title, authors, date, DOI, abstract verified. Randomized California home-computer provision; educational outcomes assessed showed no effect.

[R4] Julian Cristia; Pablo Ibarrarán; Santiago Cueto; Ana Santiago; Eugenio Severín (2017). Technology and Child Development: Evidence from the One Laptop per Child Program. American Economic Journal: Applied Economics, 9(3), 295-320.

DOI: https://doi.org/10.1257/app.20150385

Source: https://www.aeaweb.org/articles?id=10.1257/app.20150385

Verified 2026-10-16. Publisher version has 318 schools, distinct from 319 in 2012 working-paper abstract. Cite journal version only; no math/language improvement and cognitive evidence inconclusive.

[R5] Karthik Muralidharan; Abhijeet Singh; Alejandro J. Ganimian (2019). Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India. American Economic Review, 109(4), 1426-1460.

DOI: https://doi.org/10.1257/aer.20171112

Source: https://www.aeaweb.org/articles?id=10.1257/aer.20171112

Verified 2026-10-16. Publisher authors, title, DOI, date and lottery-based after-school instructional-program findings verified. Package effect, not hardware or VR effect.

[R6] Noam Angrist; Peter Bergman; Moitshepi Matsheng (2022). Experimental evidence on learning using low-tech when school is out. Nature Human Behaviour, 6, 941-950.

DOI: https://doi.org/10.1038/s41562-022-01381-z

Source: https://www.nature.com/articles/s41562-022-01381-z

Verified 2026-10-16. Publisher article/abstract: Botswana randomized SMS and phone-call interventions, combined treatment improved learning. No extrapolated effect estimate used.

[R7] Guido Makransky; Thomas S. Terkildsen; Richard E. Mayer (2019). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. Learning and Instruction,

60, 225-236.

DOI: https://doi.org/10.1016/j.learninstruc.2017.12.007

Source: https://www.sciencedirect.com/science/article/pii/S0959475217303274

Verified 2026-10-16. Publisher and University of Copenhagen institutional record agree on title/authors/DOI/date. Experimental setting not generalized as K-12 effect.

[R8] World Wide Web Consortium (2024). Web Content Accessibility Guidelines (WCAG) 2.2. W3C Recommendation, 12 December 2024.

Source: https://www.w3.org/TR/2024/REC-WCAG22-20241212/

Verified 2026-10-16. Exact dated Recommendation verified. Testable criteria; explicitly does not address every disability-related user need; automated and human evaluation distinguished.

[R9] Joshue O'Connor; Janina Sajka; Jason White; Scott Hollier; Michael Cooper (editors) (2021). XR Accessibility User Requirements. W3C Working Group Note, 25 August 2021.

Source: https://www.w3.org/TR/2021/NOTE-xaur-20210825/

Verified 2026-10-16. Official document editors/date/status verified. Explicitly not baseline requirements or W3C Membership endorsement; cited as exploratory user-needs note.

[R10] UNESCO (2019). Recommendation on Open Educational Resources (OER). UNESCO General Conference, 25 November 2019.

Source: https://www.unesco.org/en/legal-affairs/recommendation-open-educational-resources-oer

Verified 2026-10-16. Official adopted text defines OER through public domain/open license and adaptation/redistribution permissions. No presumption public webpages are licensed.

[R11] U.S. Department of Education, Privacy Technical Assistance Center (2014). Protecting Student Privacy While Using Online Educational Services: Requirements and Best Practices. U.S. Department of Education, February 2014.

Source: https://studentprivacy.ed.gov/resources/protecting-student-privacy-while-using-online-educational-services-requirements-and-best

Verified 2026-10-16. Official guidance page and original PDF verified; scope online educational services. Historic guidance is starting point, not assertion of current universal legal compliance.

[R12] Karla Hemming; Terry P. Haines; Peter J. Chilton; Alan J. Girling; Richard J. Lilford (2015). The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. BMJ, 350, h391.

DOI: https://doi.org/10.1136/bmj.h391

Source: https://www.bmj.com/content/350/bmj.h391

Verified 2026-10-16. Publisher/PubMed authors/title/DOI/date verified. Design cautions include time effects and cluster correlation; not a power guarantee.

[R13] Brantly Callaway; Pedro H. C. Sant'Anna (2021). Difference-in-Differences with multiple time periods. Journal of Econometrics, 225(2), 200-230.

DOI: https://doi.org/10.1016/j.jeconom.2020.12.001

Source: https://doi.org/10.1016/j.jeconom.2020.12.001

Verified 2026-10-16. Publisher abstract and author-maintained software references verified. Conditional parallel trends and varying treatment timing; fallback comparative study, not magical removal of selection bias.

[R14] Don Husereau; Michael Drummond; Federico Augustovski; Esther de Bekker-Grob; Andrew H. Briggs; Chris Carswell; Lisa Caulley; Nathorn Chaiyakunapruk; Dan Greenberg; Elizabeth Loder; Josephine Mauskopf; C. Daniel Mullins; Stavros Petrou; Raoh-Fang Pwu; Sophie Staniszewska; CHEERS 2022 ISPOR Good Research Practices Task Force (2022). Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 2022) statement: updated reporting guidance for health economic evaluations. BMJ, 376, e067975.

DOI: https://doi.org/10.1136/bmj-2021-067975

Source: https://www.bmj.com/content/376/bmj-2021-067975

Verified 2026-10-16. Publisher and PubMed complete author group, title, DOI/date verified. Reporting guidance originated in health; adapted transparently, no claim of educational instrument validation.

[R15] CAST (2024). Universal Design for Learning Guidelines version 3.0. CAST.

Source: https://udlguidelines.cast.org/

Verified 2026-10-16. Official guidelines site identifies version 3.0 released 30 July 2024; recommendations not treated as causal validation of this curriculum.

[R16] U.S. Department of Education, Office of Educational Technology (2024). A Call to Action for Closing the Digital Access, Design, and Use Divides: 2024 National Educational Technology Plan. U.S. Department of Education.

Source: https://www.govinfo.gov/app/details/GOVPUB-ED-PURL-gpo229250

Verified 2026-10-16. Official Government Publishing Office record and ERIC ED641164 verify institutional author, title, January 2024 and access/design/use framework.

Source and authorship notes

Original analysis and proposed research design; no completed fieldwork, empirical effect, systematic search, accepted protocol registration, ethics approval or independent human peer review is claimed.

Source selection is purposive and claim-specific. References verified against publisher, official standards, government, UNESCO, CAST or author-maintained method records on 16 October 2026. Citation does not imply endorsement by any source institution.

No stale global out-of-school count or numerical access prevalence is asserted. Historic research is dated and contextualized.

The proposed service obligation and exit protections are normative institutional/contractual recommendations, not a statutory interpretation or legal-compliance certificate.

The randomized intervention is an enhancement above an essential-service floor; urgent accessibility and access remediation is not withheld.

AI-assisted research, drafting and source checking require author accountability and independent human editorial assessment. Newly prepared article has no documented independent human review report.

Editorial methods audit resolved the initial multi-term/cohort ambiguity: final design is a one-term matched-pair randomized delayed-start comparison with a closed baseline cohort, aligned assessments, one pre-endpoint module, and no post-crossover confirmatory contrast.

Participation uses five common curricular milestones with identical goals, timing and minimum evidence requirements across arms; extra enhancement events do not inflate the denominator. Entry-to-feedback access task and baseline research-cohort denominator are common across policies. Consent/assent and baseline measurement occur before allocation.

D05 / Accountable scholarship

From Paper Agents to Accountable Student Scholarship: A Claim-Evidence-Contribution Ledger for K-12 Research

Dr. Freedom Cheteni

The Research University | Volume 1, 2026 | Edition 1.0 | D05

Original critical integrative analysis and prospective evaluation protocol

Original analyses and prospective protocols. New independent human peer review is not recorded for this edition.

DOI: not assigned to this research edition.

Read online | Author ORCID: 0009-0006-3507-0094 | Editorial record

Abstract

Paper companions can make student research easier to inspect, but fluent explanations and persistent identifiers do not establish evidential reliability or authorship. This original critical integrative analysis proposes a claim-evidence-contribution ledger linking consequential manuscript claims to version-specific sources, human intellectual contributions, software operations, and editorial decisions. Drawing on original source-evaluation and retrieval research together with scholarly infrastructure standards, it distinguishes retrieval from verification, authentication from validation, and preservation from intellectual ownership. The architecture preserves original source attribution and rights, treats companions as software rather than authors, and protects the learner's option to demonstrate equivalent scholarship without public exposure. A prospective two-stage evaluation combines a 600-item, 100-manuscript benchmark of citation fidelity, coverage, answerability, abstention, and adversarial robustness with an eight-week, 24-class feasibility study of independent source-reasoning transfer and teacher workload. All sample sizes and procedures are proposed; no benchmark results, classroom observations, or completed independent human peer review are claimed. The contribution is a testable institutional design for publication as evidence of learning rather than as a proxy for prestige.

Keywords: Designership, K-12 research, paper companions, provenance, student authorship, source evaluation, assessment, ORCID, human editorial authority

1. Publication as evidence of learning

A school research publication makes two promises: that readers can inspect an intellectual contribution, and that the named learners can answer for the work attributed to them. An elegant article, a persistent identifier, or a fluent conversational interface does not establish either promise. A student may assemble accurate sentences without understanding the inferential steps between them. Conversely, an emerging researcher may offer a valuable observation in imperfect prose. The institutional challenge is to preserve the relationship between contribution, evidence, interpretation, and revision as assistance becomes more capable.

This article proposes a claim-evidence-contribution ledger for K-12 scholarship and the paper companions that make scholarship easier to examine. A ledger is a structured, revisable account of consequential decisions, not a transcript of every keystroke. It links an article's claims to version-specific evidence and identifies the human actions and software operations through which those claims were developed. Its educational purpose is to make responsible judgment visible without equating surveillance with rigor.

The proposal draws on formative assessment's attention to evidence used in subsequent learning decisions [R2] and research showing that evaluation of digital sources can be explicitly taught [R1]. It extends those concerns to a publication environment where a companion may retrieve passages, explain terminology, and help readers interrogate an article. The original contribution is an institutional design and a falsifiable evaluation plan. No classroom trial, software benchmark, or independent human peer review is reported as completed here.

2. Method and scope of the argument

This is a critical integrative analysis with a prospective evaluation protocol. Sources were selected purposively from original technical studies, original education research, and the organizations responsible for relevant scholarly standards. Their functions differ. Technical studies motivate test conditions; education research motivates learning outcomes; standards define metadata and attribution boundaries; child-centered guidance informs governance. These materials are not interchangeable evidence of an intervention effect.

The literature search and authoritative-page verification were conducted on 16 October 2026. The selection was not a systematic review: no exhaustive search, preregistered screening, pooled estimate, or claim of comprehensive coverage is made. The synthesis asks what a school publishing system would need to preserve if its outputs were to remain inspectable after assistance, republication, revision, and changes of software provider.

The unit of analysis is the publication episode: a learner or team frames a question, encounters sources, constructs an interpretation or study, responds to criticism, and decides what may become public. Public dissemination is optional. The same intellectual standard must be achievable through a confidential school assessment, accessible document, or supervised oral defense. This distinction matters because the right to demonstrate learning should not depend on accepting permanent public exposure.

3. Four boundaries that a paper agent must respect

First, retrieval is not verification. Retrieval-augmented generation connects language generation to an external information store [R6], but retrieval alone does not establish that a cited passage supports a particular assertion. Gao and colleagues separate correctness, fluency, and citation quality in their ALCE benchmark [R3]. The proposed ledger therefore records the evidence relation itself rather than treating the existence of a citation as sufficient.

Second, inspectability is not authorship. A learner may inspect a generated explanation without having produced the contribution described by it. Authorship must follow actual intellectual participation and accountability. CRediT supplies a standardized vocabulary for describing contributions [R8]; the ledger uses that vocabulary where appropriate while retaining a plain-language account that students and families can understand.

Third, authentication is not validation. An authenticated ORCID connection establishes the connection between a record holder and an identifier through ORCID's authorization workflow [R10]. It does not establish that the person conducted an experiment, wrote an article, or reviewed another person's work. Fourth, access is not permission. A publicly readable document may lack the rights required for republication. The ledger must keep rights evidence separate from availability and from the educational merits of a source [R12].

These boundaries prevent a predictable category error: allowing the sophistication of infrastructure to stand in for the quality of the scholarship it carries. A companion should make unsupported claims easier to discover. It should not become an additional surface through which unsupported authority is manufactured.

4. The claim-evidence-contribution ledger

The proposed ledger contains five linked records. A work record identifies the article, its authors, venue, version, dates, identifiers, and disclosure statements. A source record identifies each exact input version, its creator, source location, retrieval date, license evidence, and file checksum. A claim record gives a stable identifier to a consequential proposition and classifies it as observation, source-supported assertion, interpretation, hypothesis, or recommendation. A contribution record describes who performed the relevant intellectual work. A decision record preserves the criticism received, the response, and the authority responsible for release.

The architecture adapts the distinction among entities, activities, and responsible agents in W3C PROV-DM [R7]. Its additional educational distinction is between a source relation and an ownership relation. A student can reason from an existing study without becoming its author. A software maintainer can create the retrieval interface without becoming the author of the studies it retrieves. A publisher can preserve the resulting object without acquiring the learner's intellectual contribution.

For a consequential claim, the minimum evidence relation includes a source-version identifier, a passage locator, the relevant excerpt or allowable reference to it, the asserted relation, and a human review state. Relations include supports, contradicts, limits, motivates, and does not resolve. The last category is important: an unresolved relation can be educationally productive. Where support requires several passages, the record identifies the combination and the inference connecting them. A single plausible sentence is not allowed to conceal a missing premise.

The ledger does not claim to prove truth. A checksum establishes whether bytes match a recorded object; it does not establish that the object's claims are correct. A dated teacher decision establishes who made a judgment; it does not make that judgment infallible. Records must therefore support correction, counterevidence, and appeal. Earlier decisions remain legible while superseding decisions are clearly marked. Public readers see a concise evidence trail; authorized reviewers can inspect the richer record needed to assess it.

5. A worked hypothetical example

Consider a hypothetical secondary-school team investigating shade and playground surface temperature. Its initial claim is that planting trees would reduce heat exposure at its school. A source describing a relationship in another city may motivate the project, but it does not establish the local intervention effect. The ledger classifies that claim as a hypothesis, records the source's setting and limitations, and identifies the students' contribution as question formation and study design rather than discovery of an already established local result.

The team then collects temperature measurements under a documented sampling plan. Observation records distinguish instrument readings from calculated summaries. A student who notices that shaded measurements were taken earlier in the day contributes a consequential validity challenge. The resulting revision-matching observation times or narrowing the conclusion-receives a contribution entry. This is more informative than counting paragraphs written, prompts entered, or pages in a final portfolio.

A companion asked whether the project proves that tree planting causes lower school-wide exposure should locate the design and limitations passages and explain their scope. If the manuscript reports only observational measurements, the companion must not manufacture randomized evidence. If asked for a student's home address or an unpublished participant name, it must not search private records. If asked to explain a technical term absent from the paper, it may offer a clearly labeled general explanation only in an authorized mode, separated from assertions about the paper itself.

The example is illustrative and contains no actual school observations. Its purpose is to specify the ledger's discriminations: what came from prior scholarship, what learners observed, what they inferred, what software helped express, and what a reviewer accepted as appropriately bounded.

6. Designing the course-to-publication workflow

At project opening, learners receive the assessment criteria, source-use expectations, allowed forms of assistance, publication choices, and routes for correction. Teachers select a small set of consequential checkpoints rather than requiring continuous activity capture. A practical sequence includes question justification, source appraisal, design or argument review, one substantive revision, and a final explanation of contribution. The ledger should be generated from these ordinary instructional events wherever possible.

Before public submission, the learner chooses several claims and explains why their evidence is adequate, where it is weak, and how a plausible counterexample would change the conclusion. The teacher can ask for a comparable explanation using a new passage. This creates a distinction between a polished product and transferable judgment. Digital-source evaluation instruction provides a relevant educational precedent [R1], but whether the proposed publication workflow improves transfer remains an empirical question.

The final assessment includes the artifact, the evidence trail, and a short defense. It must not reward unnecessarily extensive logging. A multilingual learner can demonstrate judgment through a supported oral explanation, translated planning notes, or accessible visual annotations. Assistance with expression must remain distinguishable from assistance that supplies the reasoning being assessed. The

relevant boundary is the intended learning outcome, established before the task, rather than a retrospective guess based on writing style.

Publication follows a separate editorial decision. A strong course performance need not imply that an article is ready for a public journal; an editorial revision need not invalidate the learning demonstrated. Journals should accept negative findings, constrained interpretations, and failed designs when the reporting is responsible. An institution that celebrates only successful-looking projects will encourage learners to hide the very revisions through which expertise becomes visible.

7. Source-grounded companions as accountable interfaces

A companion should initially operate over a bounded, versioned manuscript package. Every paper-specific answer should identify the passages on which it relies and expose a route to the original text. Interpretive summaries, direct quotations, and suggested questions require different labels. The system should disclose whether an answer draws solely from the manuscript, from a verified supplementary source, or from a general explanation that has not been established by the paper.

Unanswerability must be a normal state. The SQuAD successor introduced by Rajpurkar and colleagues deliberately includes plausible questions unsupported by the supplied paragraph [R4]. This motivates testing questions such as an unreported sample size, an unmeasured long-term outcome, or a nonexistent approval date. A good companion says what is missing and points to the relevant methods or disclosure section, rather than treating every question as an invitation to complete the record imaginatively.

The interface must also keep document content separate from system authority. A sentence inside a paper telling software to ignore its instructions is source text, not permission to change access rules. Benchmark conditions should include irrelevant passages and misleading contextual cues: Shi and colleagues demonstrated distractibility using an arithmetic benchmark with irrelevant context [R5]. That result is not a direct estimate of error in school paper companions; it supplies a reason to test this failure mode locally.

8. Authorship, identifiers, and credit without appropriation

The system should maintain separate lists for article authors, acknowledged contributors, editorial decision makers, software creators, and platform stewards. A student who originated and defended a study must not disappear because an adult edited the prose. An adult's supervisory position does not automatically justify authorship. Contribution statements should be negotiated early, revisited after substantive changes, and approved by all named human authors.

COPE's position treats AI tools as ineligible for authorship because they cannot accept the responsibilities required of authors; it also places responsibility for assisted content on human authors [R9]. Accordingly, the companion is described as software with a version, provider, documented function, and limitations. Its role can be disclosed in methods or acknowledgments and cited as software when appropriate. It is not given a human identity or an invented credential.

ORCID identifiers should be obtained through authentication, stored with the provenance of that authentication, and associated only with the actual person concerned [R10]. A publisher's founder may be credited for platform conception or software stewardship where those contributions are real. That does not justify adding every imported source article to the founder's publication list or replacing its original authors. If a separately released companion constitutes a legitimate software output, its own creators and contribution record require review before any authorized ORCID update.

Persistent identifiers likewise identify distinct objects rather than confer quality. Existing source identifiers remain attached to their original works. Relationships among manuscripts, versions, data, and software can be represented using the typed relationships supported by Crossref [R11]. Registration decisions require the relevant publisher authority; the ledger is designed to make those decisions auditable, not to substitute for them.

9. Rights, version integrity, and intellectual custody

Each source package should preserve the original title, complete author attribution, publisher or repository, date, exact source version, original identifier, and authoritative license evidence. Where a permissible copy is preserved, its checksum and file format are recorded. The extraction used by a companion receives its own checksum, extraction-method version, and passage mapping. An updated source becomes a new linked version; an extraction change is not silently represented as an unchanged original.

Creative Commons recommends attribution through title, author, source, and license, together with identifying modifications where appropriate [R12]. For the proposed journal workflow, version-specific rights review should precede preservation and public reuse. Repository metadata licenses must not be presumed to license the article itself. Excluded third-party figures, restrictive notices, or conflicting evidence require a hold or a narrower use, rather than an automatic ingestion decision.

The ledger should preserve the provenance of a license assertion as carefully as the assertion itself. It also needs a correction route when source metadata conflict with the document. Data-documentation practice supplies a useful analogy: datasheets describe how a dataset was assembled and where it should or should not be used [R16]. The proposed manuscript package applies that discipline to the inputs,

transformations, and limits of a paper companion without claiming that documentation alone resolves rights or accuracy.

10. Stage A: a prospective technical benchmark

Before classroom deployment, the proposed study would construct a held-out benchmark from 100 rights-cleared research manuscripts across disciplines and reporting formats. Six questions per manuscript would produce 600 items: three answerable questions, two plausible but unsupported questions, and one adversarial or misleading-context question. This is a proposed target, not an assembled dataset. Exact sampling rules, exclusion reasons, passage annotations, and question categories would be frozen before system comparison. A separate development set would support prompt and retrieval tuning.

Answerable questions would include an explicit fact, a methods limitation, and a bounded inference requiring more than one passage. Unsupported questions would ask for absent information or contain a plausible false premise. Adversarial items would include an irrelevant near-match source or an instruction embedded in source text. Two trained human assessors-one with information-evaluation expertise and one with relevant subject expertise-would independently label answerability, acceptable claims, and supporting passages. Disagreements would be adjudicated before outputs were scored.

The comparison would evaluate the same locked model and settings under three conditions: unrestricted generation without the manuscript package, retrieval with citations, and retrieval with the ledger's claim boundaries and abstention rules. The primary contrast is ledger-constrained retrieval versus citation-only retrieval on unsupported-answer rate. The unrestricted condition is a benchmark comparator, not a proposed student-facing service. A deterministic passage-search interface would provide an additional practical baseline. Generation parameters, retrieval index, source versions, prompts, and software dependencies would be archived. Multiple seeded runs would expose variability where the provider supports reproducibility controls.

Citation fidelity is the proportion of claim-citation links whose cited evidence actually supports the linked claim. Citation coverage is the proportion of source-verifiable claims supplied with adequate supporting evidence. Unsupported-answer rate is scored separately on unanswerable items; refusal on answerable items measures lost utility. A prespecified guardrail would require the one-sided 95% lower confidence bound for ledger-minus-citation-only answerable-item accuracy to exceed minus five percentage points; this is a proposed institutional tolerance, not an established educational threshold. Other contrasts are secondary, with multiplicity adjustment specified before testing. Assessors would be blinded to system labels and receive outputs in randomized order. Automated entailment scores may assist auditing but

cannot constitute the sole reference standard. Document-level paired bootstrap intervals would retain dependence among questions about the same paper.

11. Stage B: a prospective classroom feasibility study

Only a system meeting the technical and safeguarding gates would enter a school study. The proposed feasibility design recruits 24 teacher-class units across 12 secondary schools, with two comparable units per school and approximately 25 students per unit. Participation, sample size, access agreements, ethics review, and family consent procedures would all require confirmation before recruitment. These numbers are planning targets. No power claim or enrolled sample is implied.

Within each school, units would be randomly allocated to an eight-week ledger-supported publication sequence or an equally resourced comparison sequence using ordinary research portfolios. Different teachers would deliver the two conditions to reduce direct within-teacher crossover. Both groups would receive the same curriculum content, source-evaluation instruction, project time, assessment criteria, and access to the same permitted assistance. The intervention contrast is the structured ledger and its instructional use, not unequal access to an attractive new tool.

The primary educational outcome would be an individually completed transfer task using an unfamiliar research packet, without the companion. Twelve claims would require judgments about support, scope, and uncertainty, plus short explanations anchored to source passages. A preregistered five-level rubric, scored 0-4 per item for a maximum total of 48, would assess evidence selection, inferential restraint, and response to counterevidence. Two blinded raters would score a calibration set and a randomly sampled overlapping subset; reliability, disagreements, and adjudication would be reported.

Secondary outcomes would include article-level citation errors, revision quality, students' ability to distinguish their contribution from assistance, and teacher minutes spent on feedback and documentation. A four-week delayed task would explore retention. Surveys would assess perceived autonomy and publication pressure, but would not substitute for performance evidence. Final publication counts and page views would be descriptive institutional outcomes, never the primary indicators of learning.

12. Analysis, decision rules, and falsification

The classroom estimand is the intention-to-treat effect on the equally weighted average participating class, adjusted for class-average baseline performance and school pairing. Each teacher-class unit contributes its mean transfer score with equal weight, so larger classes do not dominate the primary estimate; a student-weighted estimate is a separate sensitivity analysis. Because assignment occurs by teacher-class unit, analysis must respect that unit. With only 12 pairs, small-sample uncertainty is material. Randomization inference over the paired assignments and a prespecified paired analysis would accompany an appropriately adjusted multilevel sensitivity model; a large number of student records cannot repair a small number of randomized units.

The protocol would specify missing-data assumptions, attrition by arm, scoring exclusions, and sensitivity analyses before outcomes are accessed. No student would be excluded for failing to complete a public article. Baseline comparisons would describe the recruited sample, not test whether randomization succeeded. A simulation using plausible cluster correlations, score variability, and missingness would determine whether a subsequent efficacy trial is feasible; this pilot would not be relabeled a definitive effectiveness test after inspection of favorable results.

Three falsification criteria distinguish the design from a promotional demonstration. First, a ledger that improves formatted attribution but fails to improve unassisted transfer has not demonstrated the proposed learning mechanism. Second, a companion that raises citation coverage while retaining unsupported claims has not demonstrated improved evidential reliability. Third, a workflow that improves scores by imposing substantial additional uncompensated teacher work has not demonstrated sustainable institutional value.

Technical release thresholds must be fixed locally before testing and reported with confidence intervals, severity classifications, and residual risks. Zero observed privacy leakage in a small synthetic-canary test does not establish zero risk. Safety failures trigger remediation and fresh testing on held-out items. Outcome families, exploratory subgroup analyses, and protocol deviations would be distinguished explicitly so that selective reporting cannot manufacture success.

13. Privacy, accessibility, and the right not to publish

UNESCO's guidance emphasizes human-centered, age-appropriate educational governance [R13], while UNICEF's 2025 guidance places children's rights, privacy, safety, inclusion, and accountability at the center of AI policy [R14]. For this proposal, those commitments become design constraints: collect the minimum evidence needed for the learning decision; separate instructional records from public artifacts; prevent public indexing of private drafts; and make disclosure understandable to the learner.

Publication consent should be separate from course enrollment and research-study participation. Students need a meaningful choice among named publication, an appropriate alternative attribution arrangement, restricted school sharing, or private assessment. An ORCID requirement cannot erase age, account-eligibility, consent, accessibility, or safety considerations. A persistent researcher identifier is a scholarly convenience, not a justification for coercive exposure.

The proposed platform would provide keyboard-accessible reading, structured headings, usable citations, accessible downloads, captioned or transcribed defenses, and a nondigital evidence route. Dictation and translation assistance should be available without requiring disclosure of disability to a public audience. Teachers need a manageable alternative when network access fails. Audit access should be role-limited and logged, with explicit retention periods and a process for correcting or removing inappropriate student material.

The NIST AI Risk Management Framework provides a voluntary structure for governing, mapping, measuring, and managing AI risk [R15]. Applying such a framework does not itself certify legal compliance or educational benefit. Institutional staff remain responsible for applicable requirements, supplier review, incident response, and changes in the population or context of use.

14. Editorial quality gates and limits

A signature student research journal should earn authority through the quality of its decisions. The first gate checks source identity, rights, attribution, and disclosure completeness. The second checks whether the question, method, evidence, and conclusion align. The third checks the learner's contribution and capacity to explain it. The fourth is an independent human editorial decision about public release, with conflicts of interest managed separately from course grading. Automated checks may identify problems; they do not become peer reviewers.

Editorial outcomes should include acceptance, revision, redirection to a different article type, and an educationally constructive decision not to publish. Reviews should identify the most consequential validity issue, request changes proportionate to the claim, and distinguish a limitation from a fatal defect. A prospective protocol should be judged as a protocol. It must not be dressed as completed primary research because the journal seeks a substantial inaugural volume.

The proposal has limits. Ledger completion can become bureaucratic, human ratings can encode disagreement, and learners can rehearse defenses without deeper understanding. Companion benchmarks may underestimate real-world prompting and multilingual variation. A well-documented source can still be wrong. These limitations motivate the proposed transfer tasks, adversarial tests, workload measures, and correction process; they are not eliminated by them.

15. Designership as accountable intellectual participation

The proposed contribution is a way to connect scholarly infrastructure to the learner's development of judgment. Publication becomes more educationally meaningful when a reader can move from a claim to its evidence, from evidence to its limits, and from an artifact to the human contribution that produced it. A paper companion is valuable when it assists those movements while preserving uncertainty and credit.

Designership, in this formulation, is the practice of designing conditions under which people can make consequential intellectual contributions and answer for them. Its success would be demonstrated by better independent reasoning, more faithful use of evidence, equitable participation, and sustainable teacher work. The ledger and the evaluation protocol offer a testable starting architecture. They do not report that those outcomes have already been achieved.

References and verification record

References retain their original authorship and publication identity. Verification dates below record source and metadata checking; they do not confer review or approval on this article.

[R1] Sarah McGrew (2020). Learning to evaluate: An intervention in civic online reasoning. Computers & Education, 145, 103711.

DOI: https://doi.org/10.1016/j.compedu.2019.103711

Source: https://www.sciencedirect.com/science/article/pii/S0360131519302647

Verified 2026-10-16. Publisher search record and original research-group page verified; original article, not a finding about the proposed ledger. Research-group evidence: https://cor.inquirygroup.org/research/learning-to-evaluate-intervention/

[R2] Paul Black; Dylan Wiliam (1998). Assessment and Classroom Learning. Assessment in Education: Principles, Policy & Practice, 5(1), 7-74.

DOI: https://doi.org/10.1080/0969595980050102

Source: https://www.tandfonline.com/doi/abs/10.1080/0969595980050102

Verified 2026-10-16. Original review article; publisher issue-year1998 verified; online platform posting2006 is not used as the original publication year.

[R3] Tianyu Gao; Howard Yen; Jiatong Yu; Danqi Chen (2023). Enabling Large Language Models to Generate Text with Citations. Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, 6465-6488.

DOI: https://doi.org/10.18653/v1/2023.emnlp-main.398

Source: https://aclanthology.org/2023.emnlp-main.398/

Verified 2026-10-16. ACL authoritative metadata and abstract verified; the article does not evaluate this school platform.

[R4] Pranav Rajpurkar; Robin Jia; Percy Liang (2018). Know What You Don’t Know: Unanswerable Questions for SQuAD. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 784-789.

DOI: https://doi.org/10.18653/v1/P18-2124

Source: https://aclanthology.org/P18-2124/

Verified 2026-10-16. ACL authoritative metadata and abstract verified; abstention benchmark motivates proposed unanswerability tests.

[R5] Freda Shi; Xinyun Chen; Kanishka Misra; Nathan Scales; David Dohan; Ed H. Chi; Nathanael Schärli; Denny Zhou (2023). Large Language Models Can Be Easily Distracted by Irrelevant Context. Proceedings of the 40th International Conference on Machine Learning, PMLR 202, 31210-31227.

Source: https://proceedings.mlr.press/v202/shi23a.html

Verified 2026-10-16. PMLR authoritative authors, title, pages, year and abstract verified; original experiment uses arithmetic problems, not school paper agents.

[R6] Patrick Lewis; Ethan Perez; Aleksandra Piktus; Fabio Petroni; Vladimir Karpukhin; Naman Goyal; Heinrich Küttler; Mike Lewis; Wen-tau Yih; Tim Rocktäschel; Sebastian Riedel; Douwe Kiela (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. Advances in Neural Information Processing Systems, 33.

Source: https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html

Verified 2026-10-16. NeurIPS proceedings author list, year and abstract verified. The proceedings citation is used; no fabricated DOI assigned.

[R7] Luc Moreau (editor); Paolo Missier (editor) (2013). PROV-DM: The PROV Data Model. W3C Recommendation, 30 April 2013.

Source: https://www.w3.org/TR/2013/REC-prov-dm-20130430/

Verified 2026-10-16. Normative recommendation version and editors verified; other contributors are listed by W3C. The educational ledger is this article’s adaptation, not a W3C-endorsed product.

[R8] National Information Standards Organization (2022). ANSI/NISO Z39.104-2022, CRediT, Contributor Roles Taxonomy. ANSI/NISO standard.

DOI: https://doi.org/10.3789/ansi.niso.z39.104-2022

Source: https://www.niso.org/publications/z39104-2022-credit

Verified 2026-10-16. Official standard record confirms 14 roles, ANSI approval 14 January 2022, publication 8 February 2022, DOI.

[R9] Committee on Publication Ethics (2023). Authorship and AI tools. COPE position statement.

DOI: https://doi.org/10.24318/cCVRZBms

Source: https://publicationethics.org/guidance/cope-position/authorship-and-ai-tools

Verified 2026-10-16. Official COPE search-indexed statement verified; live open request returned HTTP 403, so the verification is not described as full-page retrieval. Position originated 13 February 2023; later website/PDF updates do not imply new study evidence.

[R10] ORCID (n.d.). Collecting and sharing ORCID iDs. ORCID integration guidance.

Source: https://info.orcid.org/documentation/collecting-and-sharing-orcid-ids/

Verified 2026-10-16. Official page retrieved; OAuth collection, downstream authenticated metadata, limited curated unauthenticated cases, and warnings against name-only matching verified.

[R11] Crossref (n.d.). Relationships. Crossref schema documentation.

Source: https://www.crossref.org/documentation/schema-library/markup-guide-metadata-segments/relationships/

Verified 2026-10-16. Official documentation retrieved; typed relationships and data/software identifier conventions verified. Does not itself confer registration or transfer authority.

[R12] Creative Commons (n.d.). Reusing CC-Licensed Content. Creative Commons reuse guidance.

Source: https://creativecommons.org/reusing-cc-licensed-content/

Verified 2026-10-16. Official guidance retrieved; TASL attribution and source-origin practice verified. License-specific legal code governs actual reuse.

[R13] Fengchun Miao; Wayne Holmes (2023). Guidance for generative AI in education and research. UNESCO.

DOI: https://doi.org/10.54675/EWZM9535

Source: https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research

Verified 2026-10-16. Official UNESCO page and digital-library citation verified; page original date 7 September 2023 and update 16 January 2026 distinguish the 2023 publication.

[R14] UNICEF (2025). Guidance on AI and children: Version 3.0. Recommendations for AI policies and systems that uphold child rights. UNICEF, December 2025.

Source: https://www.unicef.org/innocenti/reports/policy-guidance-ai-children

Verified 2026-10-16. Official version 3.0 page retrieved; December 2025 date and child-centered requirements verified. Not an experiment or a compliance certification.

[R15] Elham Tabassi (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). National Institute of Standards and Technology, NIST AI 100-1.

DOI: https://doi.org/10.6028/NIST.AI.100-1

Source: https://www.nist.gov/publications/artificial-intelligence-risk-management-framework-ai-rmf-10

Verified 2026-10-16. Official NIST publication record confirms author, 26 January 2023 date, report number, DOI and voluntary status.

[R16] Timnit Gebru; Jamie Morgenstern; Briana Vecchione; Jennifer Wortman Vaughan; Hanna Wallach; Hal Daumé III; Kate Crawford (2021). Datasheets for datasets. Communications of the ACM, 64(12), 86-92.

DOI: https://doi.org/10.1145/3458723

Source: https://dl.acm.org/doi/10.1145/3458723

Verified 2026-10-16. ACM official indexed publication record verifies 2021 venue/DOI; author-posted arXiv 1803.09010 verifies full author list and proposal scope. The 2018 preprint date is distinct from the 2021 journal publication.

Source and authorship notes

Original conceptual architecture and prospective protocol; no primary empirical results, completed systematic review, benchmark execution, ORCID write, DOI registration, or independent human peer review is asserted.

Author intended by commissioning user: Dr. Freedom Cheteni. No Stanford affiliation, endorsement, or historical Designership reputation is inferred.

References were verified on 2026-10-16 against primary proceedings, publisher records, research-group records, or official standards/guidance. Access limitations are recorded in individual evidenceNote fields.

AI assistance was used to research sources, organize the argument, draft prose, and formulate the evaluation protocol. Human author and editors must verify the final claims, approve authorship/contributions, and conduct any claimed independent peer review.

No source content is republished here as a newly authored Cheteni study; original authors and identifiers remain intact. General normative design choices are proposals of this article rather than recommendations attributed wholesale to a cited standard.