A09 | PRIMARY-STUDY PROTOCOL AND REPRODUCIBLE METHODS RESEARCH

From Presence to Revisable Explanation: A Factorial Primary-Study Protocol for Spatial Transfer with NASA Exoplanet Data

Agathe Bertin | First author

Hasso Plattner Institute of Design; Santa Clara University

Dr. Freedom Cheteni | Second author · Principal Investigator

ORCID 0009-0006-3507-0094

Prepared 2026-10-16 | Edition 1.1 | 4,529 article words

Primary-study protocol and reproducible methods; separate human editorial record.

Keywords: Movement Thinking, spatial transfer, factorial trial, immersive learning, exoplanet data, evidence revision, accessibility

Read online / article companion / citation exports

Contribution

A prospective blocked factorial test of evidence-revision instruction across immersive and desktop offers, with an executed NASA stimulus audit and reproducible interaction-power analysis.

Original prospective study specification; executed public astronomical-file feasibility audit and artificial-data simulations; no participant outcomes.

Abstract

Immersive school environments can increase presence without establishing that learners can explain unfamiliar phenomena after the display is removed. This article converts Movement Thinking's prediction-action-revision proposition into a prospective, blocked factorial study of spatial transfer. It also reports two completed methodological investigations: a reproducible NASA Exoplanet Archive stimulus-feasibility audit and an artificial-data sensitivity analysis. The proposed study recruits one science instructor in each of 32 schools, with four course sections per instructor and a fixed baseline cohort of 24 learners per section. Within each instructor block, sections are assigned to immersive or desktop representation crossed with explicit evidence-revision or time-matched guided comparison. The primary estimand is the display-by-instruction interaction on a delayed, blinded, 20-point transfer assessment. Astronomical records supply authentic task parameters, never student outcomes. A frozen public-query response contains 6,451 default parameter rows; sequential restrictions identify 1,692 core stimulus candidates and 1,220 with the specified two-sided uncertainty information. The audit detects candidate parameter solutions, projected-separation hazards and missing eccentricities that a superficial download would conceal. Ten thousand artificial replicates per design scenario expose the cost of powering an interaction rather than a main effect. No learner has been recruited or evaluated in this work. The contribution is an inspectable route from spatial-learning theory to primary evidence: matched content, a defensible contrast, explicit model assumptions, accessible participation, original scoring instruments and reproducible methods.

1. The superintendent's question is what survives the display

A learner who can navigate a spectacular planetary scene may remain unable to explain why an eccentric orbit does not produce equal distances in equal times. A learner who studies a quiet diagram may develop that explanation well. District leaders need evidence that distinguishes those possibilities before treating visual engagement as a return on investment. The relevant public obligation concerns the durability and portability of understanding, together with the cost and accessibility of the conditions that support it.

The companion conceptual article, Beyond the Rendered Classroom, proposes that movement between prediction, action, feedback and revision is an educational mechanism requiring investigation. Cheteni's philosophical account supplies its provenance [1]. The present article contributes a different object: a primary-study protocol and completed methodological analyses that make a particular test feasible. It does not reclassify the

conceptual argument as an experiment or claim that a compelling philosophy establishes an intervention effect.

The research problem is therefore conditional. Does an explicit evidence-revision method change the advantage, disadvantage or neutrality of offering immersion when disciplinary information and learning time are held comparable? A finding that revision instruction works equally well on both displays would matter. So would an interaction favoring desktop delivery, or greater discomfort without a transfer benefit. The protocol makes these outcomes interpretable rather than defining success as whatever a technology happens to improve.

2. Theory that permits an unfavorable finding

The Cognitive Affective Model of Immersive Learning distinguishes technological features from psychological processes and identifies plausible routes through which presence and agency may support or hinder learning [2]. It provides a reason to measure those processes, not permission to substitute them for achievement. Wilson's analysis likewise distinguishes several embodied-cognition propositions rather than treating bodily participation as a single causal explanation [15]. Movement Thinking is operationalized here as a teachable sequence of commitments, not a physiological theory.

Three empirical examples sharpen the uncertainty. Madden and colleagues compared Moon-phase instruction using immersive, desktop and hands-on approaches without finding significant overall learning differences, despite stronger preference for the immersive experience [3]. Johnson-Glenberg, Bartolomea and Kalina investigated embodied instruction across desktop and immersive platforms [4]. Their work supports examining instructional design explicitly rather than treating platform as destiny. Makransky, Terkildsen and Mayer's science-laboratory comparison demonstrates that higher presence and better learning need not travel together [5]. None estimates the effects of the present exoplanet lesson.

Petersen, Petkakis and Makransky separated immersion and interactivity in a factorial study [6]. Klingenberg and colleagues tested immersion and interaction against transfer to a concrete procedure and did not find significant transfer effects of either manipulation [7]. The present design changes the instructional factor: explicit evidential revision, rather than the degree of manipulation. It also changes the disciplinary problem and delay. These differences motivate a new test; they do not erase prior null findings or entitle the proposed intervention to a favorable expectation.

3. A spatial task with inspectable intellectual work

The lesson asks learners to evaluate a proposed observation schedule for an unfamiliar planetary orbit. Students must explain what an orbital model can predict, distinguish equal angular from equal temporal intervals and revise a claim after encountering contrary evidence. The consequence is disciplinary rather than theatrical: a poorly reasoned schedule misses the observations needed to distinguish competing explanations. No claim is made that the exercise is an operational telescope-planning tool.

The unit of learning work is a documented argument. A learner states a prediction, chooses a comparison, interprets the resulting positions, identifies an alternative explanation and records whether the prediction should change. This sequence is deliberately small. It does not require a large portfolio, an elaborate publication or a public identity. The substantive question is whether the learner can justify a revision when a new constraint appears after instruction has ended.

Transfer must be specified across dimensions rather than invoked as a general virtue [8]. The delayed assessment changes the host system, viewing orientation, available evidence and decision constraint. It also removes the training display. Success therefore requires more than repeating a familiar control sequence. However, this remains disciplinary transfer within orbital reasoning. The protocol does not claim to measure general intelligence, occupational readiness across all domains or the educational superiority of immersive schooling.

4. Why the archive is a stimulus source, not an outcome database

The NASA Exoplanet Archive supplies astronomically grounded parameters and literature provenance [9]. Its Planetary Systems table has a persistent dataset identifier, 10.26133/NEA12 [10]. The table's reference-specific structure permits a more coherent starting point than combining whichever values happen to fill a row most completely. The archive documentation distinguishes that structure from the composite-parameter table, whose more complete values are not necessarily mutually consistent [11].

This distinction matters educationally. A model built from incompatible parameters can reward a learner for accepting a visually plausible inconsistency. Conversely, unexplained missing values can turn an apparently authentic lesson into an invented dataset. The present audit therefore retains solution type, discovery method, default selection, controversy and system-composition flags together with orbital parameters and the original reference field. It does not reduce provenance to a NASA logo.

The archive contains no assessment scores, classroom allocation, learner characteristics or instructional exposure for this study. Its records cannot show that students learned, nor can their scientific authority validate a pedagogy. The completed database analysis establishes a supply of candidate stimuli under declared restrictions. Human learning outcomes require separately collected primary data. Keeping these evidence layers distinct is central to the article's claim of originality and to its limits.

5. Frozen retrieval and an executed feasibility audit

The synchronous Table Access Protocol endpoint was queried using explicit selected columns, a default_flag=1 restriction and ordering by planet name [12]. The complete query and encoded URL are supplied with the analysis. The response was retrieved at 2026-10-16T00:25:01.360504+00:00, corresponding to the evening of October 15 in California. Its 1,873,252 bytes have SHA-256 034f619a7366eea09fdf2a169d9c7c9ca61f926731702f29fe9ff9e61b4b56cd. A future query may return different contents; the identifier names a living table, not this exact response.

The downloaded response contains 6,451 unique default planet-name rows. Of these, 6,378 have solution type Published Confirmed, 69 have a Kepler project-candidate solution and four have a TESS project-candidate solution. A default flag alone is consequently insufficient for the stimulus rule. These are classifications in this particular table response, not a fresh estimate of the number of confirmed planets in the universe or a claim that the archive is erroneous.

Across the unfiltered response, 368 orbital periods, 2,496 semi-major axes, 3,678 eccentricities and 4,348 inclinations are missing or nonfinite under the parser's declared check. These counts overlap and must not be added. Missing eccentricity remains missing; it is never replaced by zero to create a drawable ellipse. Inclination is audited but not required, because the lesson's viewing rotation is explicitly stipulated rather than presented as a reconstructed sky orientation.

6. Selection rules protect the meaning of a rendered orbit

The sequential restrictions retain Published Confirmed solutions, a zero controversy flag, one stellar component with no circumbinary flag, and discovery methods Transit, Radial Velocity or Transit Timing Variations. The method restriction avoids using imaging or microlensing projected separation as though it were a measured orbital semi-major axis, a documented semantic hazard [11]. It is a conservative task-design rule, not a judgment that excluded discoveries are scientifically inferior.

The remaining checks require positive finite period and semi-major axis, finite eccentricity between zero inclusive and 0.95 exclusive, zero limit flags for all three quantities and a nonempty parameter-reference field. The sequential retained counts are 6,378; 6,320; 5,761; 5,386; 3,026; 1,916; 1,692; and 1,692. These yield 1,692 core candidates, including 375 with an exact-zero eccentricity entry. Such an entry is not certified here as an independently measured perfectly circular orbit; a selected teaching case must inspect whether circularity was fixed during fitting.

A stricter uncertainty-challenge subset requires finite, strictly positive upper and strictly negative lower errors for all three parameters, a positive lower period and axis bound, and eccentricity bounds within the elliptic domain. It contains 1,220 candidates; 472 core candidates fail that additional rule. This establishes ample candidate supply without declaring every surviving record classroom-ready. An astronomer and accessibility specialist must still approve the frozen task set before recruitment. No such approval is claimed as completed.

7. Physics, visual conventions and uncertainty are separate layers

The instructional kernel solves the two-body Kepler equation $M=E-e\sin(E)$, then calculates $x=a(\cos(E)-e)$ and $y=a\sqrt{(1-e^2)}\sin(E)$. Distances are in astronomical units; period is in days; eccentricity is dimensionless. Mean anomaly advances by a stipulated fraction of the period. The executable geometry function uses a bounded bisection solver, with tests for a circular orbit, periapsis, apoapsis and invalid eccentricities. This is inspectable pedagogical geometry, not an orbit-fitting pipeline.

Time zero, viewing rotations, marker size and display scaling are task conventions. They are identified as such in every case. No actual present-day sky position, transit prediction, planetary surface, atmospheric condition or habitable status is inferred. Planet-marker radii are diagrammatic. A persistent legend explains that spatial distances and marker sizes use different scales. The lesson excludes multi-body interactions, long-term stability, light-travel effects and other complexities inappropriate to the target inference.

Reported marginal error bounds do not supply a covariance matrix or posterior distribution. The uncertainty activity therefore compares declared bounded scenarios without calling them a 95% confidence region or randomly sampling parameters as independent truths. Learners are rewarded for noticing this limitation. The candidate case bank is split by host star across training, immediate and delayed forms, preventing another planet in a familiar system from quietly serving as a supposedly unfamiliar assessment case.

8. Rights, acknowledgment and bounded redistribution

Public access is not treated as an article-wide Creative Commons license. The archive's acknowledgment and DOI instructions were checked, as was its institutional disclaimer [13]. This package redistributes original code, original instructional specifications and aggregate audit results. It does not redistribute the retrieved raw CSV, source papers, archive images, institutional marks or third-party figures. The query and response hash support transparent provenance without inventing a blanket grant of rights.

The work acknowledges the NASA Exoplanet Archive, operated by the California Institute of Technology for NASA's Exoplanet Exploration Program. Its contributing researchers remain credited through each selected parameter reference. The future classroom renderer must attach those references to its frozen case manifest and confirm any additional material-specific permissions. An archive citation does not replace attribution to the literature supplying a particular solution [13].

These boundaries also prevent an educational authorship error. Dr. Freedom Cheteni is the author of this proposed instructional and evaluation framework, not the discoverer of the planetary systems or author of their archived measurements. NASA and Caltech are data-service providers, not collaborating authors or endorsers of the protocol. The publication does not transfer astronomy papers into the journal's original research collection or place their authorship under a school leader's name.

9. The prospective factorial design

The proposed recruitment target is 32 schools, one science instructor per school, four eligible course sections per instructor and 24 consented baseline-cohort students per section: 128 sections and 3,072 learners. These are planning values, not enrollments. Schools must offer a comparable orbital-reasoning unit in grades 9-11. A pilot outside the main cohort will examine curriculum fit, accessibility and scoring before final registration. Its results may justify a transparently dated design amendment before allocation.

Consent, roster closure and baseline assessment precede randomization. An independent allocation custodian assigns the four sections within each instructor block to the four cells, using a recorded seed after roster lock. Each instructor teaches exactly one section in each condition. The public utility verifies unique sections, one instructor per school and four sections per block. A recruitment contingency does not permit silently replacing an incomplete block after outcomes are known.

Having one instructor block per school avoids assuming that several instructors inside the same school provide independent blocks. Shared instructor and school conditions are held

within each contrast. It does not eliminate interference between that instructor's sections. The study therefore documents teaching order, shared materials and cross-section discussion, while treating the no-material-spillover assumption as a genuine vulnerability. Students retain access to normal educational support throughout.

10. Four conditions with equivalent disciplinary opportunity

Factor I is the offer of an immersive stereoscopic presentation with a seated, head-tracked view, versus a desktop presentation of the same orbital kernel. Both permit the same variable changes, time steps, pauses, annotations and evidence comparisons. Both have the same narration, numerical precision, task cases and instructional duration. The display contrast inevitably includes some interface differences; it is an offer-of-mode estimand, not a pure manipulation of a disembodied quantity called immersion.

Factor R is explicit prediction-evidence-revision instruction versus a time-matched guided comparison. In R=1, a learner records a prediction before each manipulation, examines a discrepant observation and states a reason for retaining or changing the claim. In R=0, learners examine the same examples, receive the same scientific explanations and complete matched classification and summary activities. They are not denied ordinary feedback. The contrast concerns the structured evidential commitment-and-revision sequence rather than time, basic content or exposure to correct answers.

Three 45-minute lessons occur within one instructional week. A common familiarization session precedes them. Prerecorded explanations and instructor prompts limit differential coaching. Sessions are scheduled as close together as practical, teaching order is randomized within logistical constraints, and no condition receives the other condition's worksheets before the delayed endpoint. Such precautions reduce contamination but cannot certify its absence.

11. Accessibility defines the intervention being offered

Every condition includes an equivalent non-headset route, keyboard or switch-compatible control where needed, captions, adjustable text, non-color-only cues and the ability to pause. The W3C XR Accessibility User Requirements note informs this design but is not presented as a certification or mandatory technical standard [14]. The pilot must include people with relevant access needs, rather than inferring usability from a functioning device or an adult researcher's inspection.

Learners may decline or stop headset use without penalty and continue with the matched desktop route. A person who elects that route remains in the assigned immersive-offer

group for the intention-to-treat analysis. The target is consequently the effect of offering the accessible instructional package in school practice. It is not a headset-complier effect. A secondary exposure summary records actual delivery without conditioning the principal causal estimate on a post-allocation choice.

No rapid artificial locomotion, forced standing or uninterrupted exposure is required. Brief symptom checks precede and follow each lesson, with breaks and immediate stopping available. The draft symptom questions are local monitoring items, not a validated diagnostic instrument. Device errors, adverse experiences and access failures are recorded by condition. A lesson that produces excellent scores for a narrow group while excluding classmates would fail the district's implementation criterion.

12. Primary endpoint and original scoring instrument

The primary endpoint is the total score on a 35-minute assessment administered 21 calendar days after the final lesson, with an allowed window of plus or minus three school days. No headset is used. Learners receive a new host system, a changed viewing representation and an observation constraint not practiced during training. A standardized evidence update midway through the task requires a justified revision. The same task information is available in accessible visual, textual and supported response forms.

The original rubric has five dimensions, each scored zero through four: reference-frame coordination; distinction between orbital geometry and timing; evidential revision; evaluation of a rival explanation; and recognition of model or uncertainty limits. Total scores range from zero through 20. The instruments document supplies task prompts, anchors and rater instructions. This is a proposed measure requiring piloting and validity investigation. It is not labeled a validated spatial-intelligence test.

Two trained human raters independently score every de-identified response while masked to school, instructor, condition and display-use history. The prespecified endpoint is their mean total score. Raters may see the learner's initial and revised reasoning because the construct requires that comparison. A third reader investigates disagreement for calibration and reporting, without selectively replacing inconvenient primary scores. Inter-rater agreement, score distributions and masking failures are reported rather than treated as cosmetic quality assurances.

13. Measurement schedule and evidential independence

Baseline assessment measures prerequisite orbital knowledge and a short transfer form before allocation. Immediate assessment after lesson three uses a different case to measure short-term explanation. The delayed form is reserved for the primary endpoint. Form order is counterbalanced through predetermined student-level sequences within each section, and training and assessment host systems do not overlap. Pilot checks examine whether nominally parallel forms introduce avoidable difficulty differences.

Presence, perceived effort, actual time on task, prior headset use, access route and symptoms are secondary process measures. They do not form part of the transfer score. The instructional prediction-revision records describe implementation and possible mechanisms but are not rated as the independent primary outcome. This separation prevents the circular conclusion that a revision programme worked because students filled in its revision form.

Assessment administration is scripted and available to every baseline-cohort learner, including those who missed a lesson or stopped headset use. Attendance, reasons for missingness and opportunity for an accessible make-up session are recorded before outcome scoring. Teachers do not use the experimental outcome to determine course grades. Research participation, ORCID registration and public publication are separate decisions; none is a condition for receiving the lesson or educational support.

14. Estimand, clustering and the primary statistical contrast

Let μ_{bir} denote the mean delayed score for the fixed baseline cohort in school-instructor block b assigned immersion offer i and revision instruction r. The primary contrast is $\Delta_b=\mu_{b11}-\mu_{b10}-\mu_{b01}+\mu_{b00}$, averaged with equal block weights. It asks whether the effect of revision instruction differs between display offers. The direction is not presupposed. Revision and immersion marginal effects are declared secondary contrasts, each averaging its two simple effects.

With one section per factorial cell per block, the sample variance of block contrasts divided by the number of blocks supplies a conservative across-block uncertainty estimate under independent blocks and the stated no-interference conditions. The principal interval and two-sided test use an approximate t reference with $B-1$ degrees of freedom. This is not an exact small-sample guarantee. Heterogeneous block effects contribute to the estimated variability; common school and instructor intercepts cancel within each complete contrast.

The analysis specification defines section means, equal weights, identifiers and reporting conventions, following the need to make cluster allocation and analysis units explicit [16]. A learner-level secondary model may improve precision using baseline performance and form sequence, with the nesting preserved. It does not replace the prespecified block contrast after results are seen. Permuting observed outcomes does not give an exact test of zero interaction when unknown main effects remain; no such claim is made by the supplied tools.

15. Missingness, multiplicity and departures from the plan

The intention-to-treat population is the fixed baseline cohort. The executable complete-cohort estimator deliberately refuses missing scores or incomplete blocks; it is a verification tool, not a covert listwise-deletion policy. The planned full analysis uses multilevel multiple imputation for missing delayed outcomes, with 50 imputations and predictors including baseline score, form, allocation cell, attendance, immediate assessment, access route and reported symptoms. School-block and section effects preserve clustering. Bounded predictive draws keep scores within zero through 20.

The supplementary specification declares the imputation model, convergence checks, Barnard-Rubin small-sample pooling [20] and missing-not-at-random sensitivity shifts. Those procedures are specified rather than executed because there are no participant outcomes to impute. If an entire section is missing, reporting must show the resulting dependence on extrapolation. Worst-case bounded outcomes accompany model-based conclusions. A visibly precise interval cannot make informative dropout ignorable.

There is one confirmatory interaction endpoint. The two marginal effects form a secondary family with Holm adjustment; immediate transfer and process variables are reported with estimates and clearly exploratory intervals. Associations between revision traces and delayed scores are not identified causal mediation effects. Any changes to recruitment, scoring, delivery or the endpoint window require versioned reasons, dates and visibility into whether outcome information was available when the decision occurred.

16. Completed design sensitivity: interactions are expensive

The executed simulation uses NumPy 2.3.5, seed 20261016 and 10,000 artificial replicates for each of nine scenarios. It varies instructor-school blocks across 16, 32 and 48 and section-level intracluster correlation across 0.05, 0.10 and 0.15, with 24 learners per section. The working model has unit marginal individual variance, independent section residuals within a block, complete outcomes and a constant interaction of 0.40 standard deviations. Common block effects cancel by construction. Cluster-design sensitivity is a planning requirement rather than a reassurance to be supplied after recruitment [17].

For the proposed 32 blocks, simulated interaction power is 0.9538, 0.8457 and 0.7187 as the assumed correlation increases. The corresponding Monte Carlo standard errors are 0.00210, 0.00361 and 0.00450. Null rejection proportions are 0.0469, 0.0500 and 0.0462. These near-nominal artificial-data results check implementation under the simulated model; they do not establish operating characteristics under contamination, attrition or skewed bounded scores.

At 48 blocks, power under the highest assumed correlation is 0.8854. The 32-block normal-approximation interaction thresholds for 80% power are approximately 0.296, 0.367 and 0.427 standard deviations. A 0.40-standard-deviation interaction is a scenario, not a predicted benefit or validated importance threshold. The Gaussian calculations omit floor and ceiling effects. Pilot information and resources must determine whether a larger prospectively fixed design is necessary before main-study allocation.

17. What would count against the account

The strongest theoretical result is not necessarily a positive interaction. An accurately estimated near-zero interaction with a beneficial marginal revision effect would suggest that the evidential method travels across media. A negative interaction could indicate that the method competes with immersive interface demands. An immersive presence advantage accompanied by no delayed-transfer benefit would challenge the procurement rationale while remaining consistent with prior distinctions between experience and learning [2] [3] [7].

The mechanism would be weakened if students produced more elaborate revision records without better independent reasoning, or if gains appeared only on a familiar training representation. A benefit restricted to experienced headset users would prompt an access and instructional-design investigation, not a declaration that novices lack aptitude. Excess symptoms, loss of instructional time and exclusion from the accessible task are implementation harms even when a mean score favors a condition.

No equivalence claim follows from a nonsignificant test. Before recruitment, investigators and affected educators must choose and justify any smallest educationally important difference in rubric points. Cost per delivered accessible lesson and time needed for teacher preparation are reported alongside learning. The public decision is whether a defensible benefit justifies the whole implementation burden, rather than whether a device can produce one attractive statistic.

18. Ethics, governance and a study that has not begun

This protocol is prepared for prospective registration and independent ethics review. No registration number, ethics approval, participating school, learner consent, recruitment total or trial outcome is asserted. The structure is informed by SPIRIT's emphasis on transparent prospective specification while adapting it to educational research rather than claiming biomedical protocol compliance [18]. Schools and families must be able to distinguish the existence of a publication from authorization to conduct a study.

Future primary records require pseudonymous identifiers, restricted access, a retention schedule and separate storage of consent and outcome data. Public sharing should favor code, scoring guidance and appropriately screened aggregate results. Student handwriting, voice, room imagery and behavioral telemetry are not routine public supplements. A data-access committee must evaluate re-identification risk before any individual-level release, with the authority to refuse publication of identifying artifacts.

The second author and Principal Investigator is also Editor in Chief. A future acceptance decision for this new protocol and any results paper requires documented independent editorial handling and human expertise. The companion conceptual article's earlier review does not constitute review of this new design. This edition records its actual state without inventing an acceptance date, DOI assignment, institutional collaboration or reviewer endorsement.

19. Limitations and the distinct contribution

Database completeness favors scientifically well-characterized systems and selected discovery methods; the stimulus candidates are not representative of all exoplanets. The strict uncertainty rule may omit useful cases, while published marginal errors omit covariance and other sources of uncertainty. The audit identifies possibilities, not a finished validated curriculum. A selected model may remain unsuitable for novice learning despite passing every numerical check.

The primary study would evaluate one topic, age range, instructional sequence and accessible delivery package. Instructor spillover, school recruitment, technology familiarity, assessment validity and missing outcomes can alter interpretation. The four-section requirement may select unusually large courses or schools. The simulation supplies no empirical estimate of the study's actual correlation or likely effect. Generalization to other sciences, adult employment or an entire university would require additional studies.

The contribution is narrower and more durable: it connects a philosophical learning proposition to a falsifiable comparison, gives an authentic database a precisely limited role, and exposes assumptions before human outcomes exist. A district can inspect why a case is eligible, how a lesson changes a learner's intellectual obligations, which score will matter and what evidence would justify changing course. That is an infrastructure for primary research rather than a claim to have completed a field experiment.

Method and data statement

Completed work consists of the frozen NASA query-response audit, nine-scenario artificial-data sensitivity analysis, geometry checks, estimator checks and preparation of the original protocol and instruments. The supplementary code can fetch a fresh public response, audit a supplied local response, generate the design scenarios, allocate a supplied roster and analyze a complete synthetic or study cohort under declared restrictions. The original retrieved CSV is not included. Its query, retrieval timestamp and hash are retained, so a later response must not be misrepresented as the same snapshot.

There are zero actual trial participants in this package. The multiple-imputation model, optional adjusted models and curriculum validation remain prospective specifications. The archive provides astronomical task parameters only. Full-text verification, abstract or metadata verification, official-document inspection and actual file analysis are distinguished in the source audit. A reference's inclusion does not assert that its original full text, images or data have been relicensed for republication.

AI, authorship and editorial disclosure

ChatGPT assisted research planning, source discovery, drafting and code preparation [19]. The supplied programs executed the stated audit and artificial-data calculations; the assistant did not generate participant observations or serve as a human peer reviewer. The exact source-model build is not asserted. The paper agent retrieves passages from this manuscript and must not describe proposed results as observed effects or extend its sources beyond their documented evidential role.

Agathe Bertin is the first author, with affiliations at Hasso Plattner Institute of Design and Santa Clara University. Dr. Freedom Cheteni is the second author and Principal Investigator of this original framework and protocol. Final author approval, independent review, study authorization and any later acceptance are distinct records. NASA, Caltech, the cited researchers and the conceptual article's reviewer are not made coauthors or endorsers by citation. Original astronomers retain credit for source measurements, and any future results article must identify actual investigators, contributions, deviations and permissions before publication.

Research materials and reproducibility

Inspect the prospective study specification, original draft instruments, executed public-file audit, design simulations and documented tools. No local participant trial outcomes are reported.

Complete research package · ZIP

Research methods guide

Study specification

Task instruments and scoring

Statistical analysis specification

Executed NASA stimulus feasibility audit

Design simulations and uncertainty

Source and rights audit

Executable research tools

Methods tests

Read the corresponding conceptual foundation

Evidence and limits

Claim	Evidence	Limit
Authentic astronomical stimuli can be selected under explicit semantic restrictions.	Executed 6,451-row default-solution audit; 1,692 core and 1,220 strict-uncertainty candidates. [9, 10, 11, 12]	Numerical eligibility is not astronomy/curriculum approval or learning evidence.
Presence and delayed transfer require distinct measures.	Prior empirical platform and transfer studies motivate independent assessment. [2, 3, 5, 7]	Prior studies do not establish effectiveness of the proposed lesson.
A display-by-method interaction is a distinct primary research target.	Blocked 2×2 allocation, equal disciplinary opportunity, delayed blinded score and explicit contrast. [6, 16, 18]	The study is prospective and relies on limited cross-section interference.
Powering an interaction is sensitive to the number of independent school blocks.	Nine executed 10,000-replicate artificial-data scenarios with Monte Carlo uncertainty. [17]	No attrition, floor/ceiling, treatment heterogeneity or spillover is modeled.

References and rights

[1] Freedom Cheteni. Movement Thinking: The Human World Model. The VR School · Media. 2026-08-03.

https://www.thevrschool.org/media/world-models-and-learning

Citation only; no article-wide reuse license asserted. Original authors and institutions retain their rights; source text, figures and raw data are not republished.

Authoritative source / evidence | checked 2026-10-16

[2] Guido Makransky; Gustav B. Petersen. The Cognitive Affective Model of Immersive Learning (CAMIL): a Theoretical Research-Based Model of Learning in Immersive Virtual Reality. Educational Psychology Review, 33, 937-958. 2021-01-06.

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

CC BY 4.0; citation only in this work. Original authors and institutions retain their rights; source text, figures and raw data are not republished.

Authoritative source / evidence | checked 2026-10-16

[3] J. Madden; S. Pandita; J. P. Schuldt; B. Kim; A. S. Won; N. G. Holmes. Ready student one: Exploring the predictors of student learning in virtual reality. PLOS ONE, 15(3), e0229788. 2020-03-25.

https://doi.org/10.1371/journal.pone.0229788

CC BY 4.0; citation only in this work. Original authors and institutions retain their rights; source text, figures and raw data are not republished.

Authoritative source / evidence | checked 2026-10-16

[4] Mina C. Johnson-Glenberg; Hannah Bartolomea; Elena Kalina. Platform is not destiny: Embodied learning effects comparing 2D desktop to 3D virtual reality STEM experiences. Journal of Computer Assisted Learning, 37(5), 1263-1284. 2021-06-01 (online).

https://doi.org/10.1111/jcal.12567

Citation only; no article-wide reuse license asserted. Original authors and institutions retain their rights; source text, figures and raw data are not republished.

Authoritative source / evidence | checked 2026-10-16

[5] Guido Makransky; Thomas Schjødt Terkildsen; Richard E. Mayer. Adding immersive virtual reality to a science lab simulation causes more presence but less learning. Learning and Instruction, 60, 225-236. 2019-04-01.

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

Citation only; no article-wide reuse license asserted. Original authors and institutions retain their rights; source text, figures and raw data are not republished.

Authoritative source / evidence | checked 2026-10-16

[6] Gustav Bøg Lassen Petersen; Giorgos Petkakis; Guido Makransky. A study of how immersion and interactivity drive VR learning. Computers & Education, 179, 104429. 2022-04.

https://doi.org/10.1016/j.compedu.2021.104429

Citation only; no article-wide reuse license asserted. Original authors and institutions retain their rights; source text, figures and raw data are not republished.

Authoritative source / evidence | checked 2026-10-16

[7] Sara Klingenberg; Robin Bosse; Richard E. Mayer; Guido Makransky. Does Embodiment in Virtual Reality Boost Learning Transfer? Testing an Immersion-Interactivity Framework. Educational Psychology Review, 36, 116. 2024-10-04.

https://doi.org/10.1007/s10648-024-09956-0

Citation only; no article-wide reuse license asserted. Original authors and institutions retain their rights; source text, figures and raw data are not republished.

Authoritative source / evidence | checked 2026-10-16

[8] Susan M. Barnett; Stephen J. Ceci. When and where do we apply what we learn? A taxonomy for far transfer. Psychological Bulletin, 128(4), 612-637. 2002-07.

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

Citation only; no article-wide reuse license asserted. Original authors and institutions retain their rights; source text, figures and raw data are not republished.

Authoritative source / evidence | checked 2026-10-16

[9] Jessie L. Christiansen; Douglas L. McElroy; Marcy Harbut; David R. Ciardi; Megan Crane; John Good; Kevin K. Hardegree-Ullman; Aurora Y. Kesseli; Michael B. Lund; Meca Lynn; Ananda Muthiar; Ricky Nilsson; Toba Oluyide; Michael Papin; Amalia Rivera; Melanie Swain; Nicholas D. Susemiehl; Raymond Tam; Julian van Eyken; Charles Beichman. The NASA Exoplanet Archive and Exoplanet Follow-up Observing Program: Data, Tools, and Usage. The Planetary Science Journal, 6, 186. 2025-08.

https://doi.org/10.3847/PSJ/ade3c2

Citation only; no article-wide reuse license asserted. Original authors and institutions retain their rights; source text, figures and raw data are not republished.

Authoritative source / evidence | checked 2026-10-16

[10] NASA Exoplanet Science Institute. NASA Exoplanet Archive: Planetary Systems Table. Caltech/IPAC, NASA Exoplanet Archive. Living dataset; snapshot retrieved 2026-10-16.

https://doi.org/10.26133/NEA12

Citation only; no article-wide reuse license asserted. Original authors and institutions retain their rights; source text, figures and raw data are not republished.

Authoritative source / evidence | checked 2026-10-16

[11] NASA Exoplanet Archive. Planetary Systems and Planetary Systems Composite Parameters Table Definitions. Caltech/IPAC, NASA Exoplanet Archive. Accessed 2026-10-16.

https://exoplanetarchive.ipac.caltech.edu/docs/API_PS_columns.html

Citation only; no article-wide reuse license asserted. Original authors and institutions retain their rights; source text, figures and raw data are not republished.

Authoritative source / evidence | checked 2026-10-16

[12] NASA Exoplanet Archive. Retrieving Exoplanet Archive Data with Table Access Protocol. Caltech/IPAC, NASA Exoplanet Archive. Accessed 2026-10-16.

https://exoplanetarchive.ipac.caltech.edu/docs/TAP/usingTAP.html

Citation only; no article-wide reuse license asserted. Original authors and institutions retain their rights; source text, figures and raw data are not republished.

Authoritative source / evidence | checked 2026-10-16

[13] NASA Exoplanet Archive. Acknowledging the NASA Exoplanet Archive in Publications; Exoplanet Archive Privacy Policy and Disclaimers. Caltech/IPAC, NASA Exoplanet Archive. Acknowledgment page updated 2025-08-13; accessed 2026-10-16.

https://exoplanetarchive.ipac.caltech.edu/docs/acknowledge.html

Citation only; no article-wide reuse license asserted. Original authors and institutions retain their rights; source text, figures and raw data are not republished.

Authoritative source / evidence | checked 2026-10-16

[14] Joshue O’Connor; Janina Sajka; Jason White; Scott Hollier; Michael Cooper (editors). XR Accessibility User Requirements. World Wide Web Consortium, Accessible Platform Architectures Working Group Note. 2021-08-25.

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

Citation only; no article-wide reuse license asserted. Original authors and institutions retain their rights; source text, figures and raw data are not republished.

Authoritative source / evidence | checked 2026-10-16

[15] Margaret Wilson. Six views of embodied cognition. Psychonomic Bulletin & Review, 9, 625-636. 2002-12.

https://doi.org/10.3758/BF03196322

Citation only; no article-wide reuse license asserted. Original authors and institutions retain their rights; source text, figures and raw data are not republished.

Authoritative source / evidence | checked 2026-10-16

[16] Marion K. Campbell; Gilda Piaggio; Diana R. Elbourne; Douglas G. Altman; for the CONSORT Group. Consort 2010 statement: extension to cluster randomised trials. BMJ, 345, e5661. 2012-09-04.

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

Citation only; no third-party text or instrument republished. Original authors and institutions retain their rights. Data and instrument use must follow the source terms.

Authoritative source / evidence | checked 2026-10-16

[17] Clare Rutterford; Andrew Copas; Sandra Eldridge. Methods for sample size determination in cluster randomized trials. International Journal of Epidemiology, 44(3), 1051-1067. 2015.

https://doi.org/10.1093/ije/dyv113

Citation only; no third-party text or instrument republished. Original authors and institutions retain their rights. Data and instrument use must follow the source terms.

Authoritative source / evidence | checked 2026-10-16

[18] An-Wen Chan; Isabelle Boutron; Sally Hopewell; David Moher; Kenneth F Schulz; Gary S Collins; Ruth Tunn; Rakesh Aggarwal; Michael Berkwits; Jesse A Berlin; Nita Bhandari; Nancy J Butcher; Marion K Campbell; Runcie C W Chidebe; Diana R Elbourne; Andrew J Farmer; Dean A Fergusson; Robert M Golub; Steven N Goodman; Tammy C Hoffmann; John P A Ioannidis; Brennan C Kahan; Rachel L Knowles; Sarah E Lamb; Steff Lewis; Elizabeth Loder; Martin Offringa; Philippe Ravaud; Dawn P Richards; Frank W Rockhold; David L Schriger; Nandi L Siegfried; Sophie Staniszewska; Rod S Taylor; Lehana Thabane; David J Torgerson; Sunita Vohra; Ian R White; Asbjørn Hróbjartsson. SPIRIT 2025 statement: updated guideline for protocols of randomised trials. BMJ, 389, e081477. 2025-04-28.

https://doi.org/10.1136/bmj-2024-081477

Citation only; no third-party text or instrument republished. Original authors and institutions retain their rights. Data and instrument use must follow the source terms.

Authoritative source / evidence | checked 2026-10-16

[19] OpenAI. ChatGPT [AI-assisted research planning, drafting and code preparation software]. Software/tool disclosure; not a human author or peer reviewer. Accessed 2026-10-15-16.

https://chatgpt.com/

Tool use disclosed; model version not asserted. Original authors and institutions retain their rights; source text, figures and raw data are not republished.

Authoritative source / evidence | checked 2026-10-16

[20] John Barnard; Donald B. Rubin. Small-sample degrees of freedom with multiple imputation. Biometrika, 86(4), 948-955. 1999.

https://doi.org/10.1093/biomet/86.4.948

Citation only; no third-party text or instrument republished. Original authors and institutions retain their rights. Data and instrument use must follow the source terms.

Authoritative source / evidence | checked 2026-10-16

An original educational primary-study protocol and methods analysis. Astronomical parameters retain their original source attribution. NASA/Caltech provide a public data service and are not collaborators or endorsers. Source articles, images and the raw TAP response are not republished. New work licensing and final human author approval are separate records.