Traceable AI instruction for pilot judgment

Delayed learning and transfer to simulated flight planning

RESEARCH DESIGN MANUSCRIPT

Prepared for Ahmed Allem’s review · Aviation Practice

Abstract

AI-assisted ground school can improve the quality of an answer while the tutor is visible. The educational question is what learners can still explain once that support is removed. This dissertation manuscript proposes a civil-aviation research program linking source-grounded feedback to delayed reasoning and performance on unfamiliar simulated planning tasks.

The proposed Study 1 is an offline paired assessment of a frozen Aviation Infinity tutor against instructor-designed feedback. Independent raters score factual support, scope, and appropriate escalation. Study 2 is a randomized comparison that holds the learning interface and content constant, with day-14 performance as the primary outcome. Study 3 uses the same assigned groups in a held-out Avioyx or neutral sandbox without the tutor.

The contribution is a research design, a traceable scoring method, and a reproducible prototype bank. The open-source-compatible companion contains 72 original fictional cases and a standard-library scoring fixture. This manuscript proposes methods and interpretation rules. Empirical chapters await approved data collection and independently checked analysis.

KEYWORDS

Aviation education · AI tutoring · delayed learning · human factors · simulation transfer · reproducibility

Status, authorship, and use

Research design manuscript prepared for Ahmed Allem’s review, 16 November 2026. Its design builds on his public portfolio and published research frameworks. It records proposed procedures, analytical choices, and prototype instruments. All fictional examples and score fixtures are labeled.

This edition was developed with AI assistance. It does not claim Allem approved or authored this draft. He should verify the literature, revise the argument and methods, add his own scholarly work, and disclose the AI assistance he uses in the version he submits. Committee decisions, ethics determinations, and participant results must be documented from their respective records.

Christian K. Jaedicke’s full Embry-Riddle dissertation was reviewed from the supplied copy. Its public catalog is linked in the reference register. The original’s introduction, methods, results reporting, and supporting appendices inform the structure for this distinct Aviation Practice study.

> **DOCUMENT STRUCTURE**
>
> Pages 1-80 contain front matter, chapters and references. Pages 81-200 contain the fictional task bank and supporting instruments.

Contents

Abstract	002
Status, authorship, and use	003
Research at a glance	005
Terminology and reading guide	006
1. Problem, purpose, and contribution	007
2. Literature and dissertation precedent	015
3. Portfolio and research infrastructure	023
4. Study 1. Offline instructional reliability	031
5. Study 2. Controlled delayed learning	039
6. Study 3. Transfer to a separate simulator task	047
7. Research integrity, accessibility, and AI	055
8. Synthesis, replication, and defense	063
References and source notes	071
A. Fictional assessment case bank	081
B. Fieldwork and committee instruments	153
C. Data dictionary and replication records	177
D. Worked verification exercises	189

Research at a glance

The three proposed studies answer three distinct questions. The first asks whether the intervention is reliable enough to use in a controlled lesson. The second asks what participants learn. The third asks whether that learning appears in a changed setting.

01 / VALIDATE

Freeze the tutor and compare it with an instructor-authored benchmark.

02 / ESTIMATE

Randomize feedback within the same lesson and measure day-14 reasoning.

03 / TEST TRANSFER

Assess unfamiliar planning tasks without the study tutor.

FIGURE 1. EVIDENCE SEQUENCE

The reliability decision precedes the learner trial. Both learner assessments use the same randomized groups. A change in interface tests whether the observed reasoning is portable.

Terminology and reading guide

TRACEABLE FEEDBACK

Guidance that identifies the supporting source, its scope, the assumption under discussion, and what remains unresolved.

DELAYED JUDGMENT

A reasoned explanation scored after an interval on an unfamiliar task under identical reference access.

TRANSFER

Use of the same underlying reasoning in a second assessment environment. The new interface changes the context while the intended construct stays fixed.

BENCHMARK AND HOLDOUT

A benchmark is a fixed comparison. A holdout is a deliberately protected set of examples for evaluating a system or scoring method after development.

> **READER NAVIGATION**
>
> PDF bookmarks and the clickable contents lead to each chapter and appendix. Source links in the references open the originating record.

The learning that survives the interface

A modern ground-school interface can make a difficult decision look simple. It can retrieve the correct reference, label uncertainty, and prompt the learner to consult an instructor. The same design can hide whether the learner understands which reference matters and why. If the software supplies the move that the assessment rewards, apparent competence may disappear outside the interface. This dissertation treats that difference as a measurable property of instruction. It begins with a support system and ends with an unaided task. [R01, R05, R14]

The central outcome is a defensible decision explanation after a delay. In a fictional planning exercise, learners must identify what is unresolved, connect the issue to a source, and explain why their proposed next check follows. This focus permits a direct study of judgment while keeping the work entirely within supervised educational settings.

> **PROTOCOL APPLICATION**
>
> Figure 1 outlines three evidentiary steps. A favorable pilot score cannot replace a documented reliability gate.

Statement of the problem

Training platforms routinely capture score, completion, and engagement. These measures answer different questions. Someone can complete many exercises, feel confident, and still fail to recognize a relevant discrepancy when the layout, airport, or vocabulary changes. The validity threat is particularly acute when both teaching and testing use the same interface or model. Shared surface cues can inflate apparent transfer.

The design addresses this with a chain of three studies. The first evaluates the tutor itself, the second estimates what learners retain under controlled conditions, and the third checks whether the same learners can reason in a different simulated interface. Each stage has a separate output and an explicit decision rule. A negative result at one stage remains interpretable evidence for the next design iteration.

> **PROTOCOL APPLICATION**
>
> Freeze a common 2D interface for the main experiment. Test independent handling of familiar concepts through new record combinations.

Research question and hypotheses

Among consenting adult student pilots at participating providers, does source-grounded AI feedback improve a day-14 reasoning score compared with equally timed, instructor-designed static feedback in the same 2D lesson? The primary hypothesis concerns an assignment effect on the observed rubric score. There is no directional confirmatory claim about timing or workload. The secondary hypothesis concerns performance in a held-out planning activity where access to the study tutor is removed.

Study 1 asks a different question about whether the frozen tutor is sufficiently reliable to justify a controlled trial. The analysis must distinguish intervention failure, failure to implement as specified, and uncertainty due to imprecision. Before recruitment, the research team fixes a minimum educationally important difference with an independent instructor panel.

> **PROTOCOL APPLICATION**
>
> The confirmatory comparison uses one primary endpoint. Time, confidence, workload, and transfer remain distinct secondary measures.

Operational definition of pilot judgment

Pilot judgment is defined narrowly here as the ability to recognize a planning discrepancy, explain the assumption it threatens, trace a source, and choose a defensible next verification action. It excludes aircraft handling and actual in-flight decisions. Tasks use fictional materials. The operational definition permits scoring without evaluating a participant’s professional entitlement or giving operational advice.

The assessment contains four separately rated dimensions. Each dimension receives zero for absent or contradicted reasoning, one for a partially supported answer, and two for a sufficiently supported answer. A plausible answer that fails on source traceability earns credit for the reasoning it actually shows. This scoring principle protects the distinction between knowing a conclusion and knowing why the conclusion follows.

> **PROTOCOL APPLICATION**
>
> A complete answer identifies the discrepancy, states its implication, names an applicable source, and proposes a next check.

Distinct contributions to knowledge

The first contribution is a construct and instrument: a design for evaluating traceability, uncertainty, and human referral in civil aviation education. The second contribution is a causal question about delayed performance after exposure to adaptive rather than static feedback. The third contribution is a portability check in a second interface. The research links the three through identifiable versions and scoring rules.

The novelty is the study design and its testable combination of methods. Individual pieces have precedents in tutoring research, simulator transfer, human-automation studies, and source-grounded reasoning. The proposed studies cannot establish that this is the first approach of its kind. The literature and evidence mapping explains the contribution in narrower, verifiable terms. [R01, R07, R12, R15]

> **PROTOCOL APPLICATION**
>
> The public benchmark provides a reproducible instrument design. Faculty review and rater validation determine whether it is suitable for field use.

What counts as meaningful transfer

Transfer changes one or more conditions between instruction and assessment. Here the theory assessment changes the fictional problem while preserving the construct. The planning assessment also changes the interface and information layout. The transfer taxonomy in Barnett and Ceci helps specify those changes rather than treating transfer as a universal distance. A broader claim would require evidence across tasks, providers, and contexts. [R08]

The same source discipline applies across both settings. A learner should trace a decision to an available record, explain what makes the record applicable, and recognize when additional confirmation is needed. The simulator assessment removes tutor access. This makes it possible to distinguish an improved supported performance from judgment demonstrated independently.

> **PROTOCOL APPLICATION**
>
> Compare instructional and assessment scenarios by source dependencies, task demands, vocabulary, and interface cues.

Intended readers and decision uses

Instructors can use the proposed scoring instrument to review the clarity of their own feedback and scenario design. Researchers can extend the task bank or test the same causal question under another delivery mode. Product teams can use the reliability gate to find which failure modes require review before deploying changes in educational environments. These decisions require different levels of evidence.

A study sponsor may find a positive result commercially attractive. The version record therefore ties the dissertation to specific study materials, and an independent reviewer signs the analysis and release decision. That creates room for null, adverse, and equivocal findings to be recorded without implying that the platform as a whole succeeded or failed.

> **PROTOCOL APPLICATION**
>
> The field roles and approvals appear in Appendix B. They help separate research design, implementation, and evidence ownership.

Scope, status, and boundaries

This manuscript is a research design draft prepared for Ahmed Allem’s review. It develops an original proposal and open-source-compatible benchmark. Its evidence chapters specify intended outputs and mock reporting structures. Credit, dissertation approval, ethics determination, recruitment commitments, instrument validation, and completed study results belong to future verified records.

Every fictional task is an educational instrument-development example. The scoring targets judgment about source use and uncertainty in a scenario. The project would require input from independent instructors before real participants use the materials. Public portfolio language describes the author’s product context. Reported adoption numbers, marketing labels, and features have no evidentiary role in estimates of learning.

> **PROTOCOL APPLICATION**
>
> The proposal sets outcomes and decisions in advance. The data dictionary distinguishes expected evidence from actual records.

Search strategy and coverage

The review begins with Stanford’s public dissertation and laboratory records, MIT’s repository, and Embry-Riddle’s Scholarly Commons. Searches combined civil aviation, flight training, simulator transfer, feedback, automation, and validation terms, then followed repository and author links. Only records with identifiable provenance were retained as principal precedents. Articles supplement the dissertation records where they answer different questions. [R01-R08, R11-R15, R21]

The supplied full Embry-Riddle dissertation and its public catalog were checked together. The review examined the abstract, the structure of chapters and appendices, and the discussion of limitations. Stanford’s direct catalog enumeration was unavailable. A later search can refresh the queries and document screening decisions using the source register.

PROTOCOL APPLICATION

Capture source access level, title, publication year, and relationship to the planned design for every item in the review.

Embry-Riddle structural model

Jaedicke’s study motivates the present dissertation’s organization through its explicit mapping from a proposed training mechanism to a manipulated feature, a defined task, measured outcomes, and bounded interpretation. Its four-condition factorial study varied haptic feedback and cockpit complexity. It measured response time and position error after standardized training. The discussion treats the null performance findings and greater perceived immersion as evidence with different meanings. [R01, Chapter V]

This drafting project takes the logic of that structure into a new question. A tool feature needs a task whose cognitive demands make the feature relevant. The report should preserve differences between subjective experience and scored performance. The present experiment therefore holds presentation constant at first. Chapter III of the example also documents orientation, practice, apparatus, a pilot test, and measurement checks. Those functions recur in Chapters 4-6 and Appendices A-C here.

PROTOCOL APPLICATION

The relevant structural precedent is a defined comparison with an observable outcome and a cautious interpretation of uncertain estimates.

Transfer in simulation

Zhang’s dissertation offers a close example of training in one simulation condition and evaluating performance on a separate flight training device. Its abstract describes a comparison between VR and conventional PC-based simulation with a specific post-training assessment. That design clarifies why a second assessment context is necessary. It also shows how transfer claims remain coupled to the moves and conditions actually tested. [R02]

Allem’s study differs in the construct it measures. The primary comparison is feedback in desktop theory learning. The additional simulation measures the handling of planning discrepancies and documentation uncertainty. A change in measured maneuver performance would answer a different question. The proposed battery therefore avoids conflating procedural execution with explanatory judgment.

> **PROTOCOL APPLICATION**
>
> Assess simulator transfer through a second context where interface cues cannot supply the intended reasoning steps.

Acceptance and meaningful use

Fussell’s dissertation considers factors associated with students’ intentions to use VR in flight training. The question matters to adoption because accessibility, comfort, perceived usefulness, and availability affect who engages. An adoption study can explain why a learning opportunity is attractive or hard to access. [R03]

The proposed trial records similar implementation factors without placing them in the primary outcome. Engagement is a possible mediator. A favorable satisfaction rating can coexist with no detectable delayed learning. The research report will present experience outcomes separately and describe barriers by condition, with suppression or grouping when counts are too small to protect privacy.

> **PROTOCOL APPLICATION**
>
> A separate satisfaction question is useful because adoption and demonstrated judgment are different outcomes.

Automation, workload, and independent oversight

Johnson’s MIT dissertation treats workload and attention as distinct from task performance in changes of automation responsibility. The context differs from tutoring, but the methodological separation transfers. Adding an AI coach can alter what someone chooses to attend to regardless of whether their observed answer improves. [R04]

The proposed trial keeps instruction time and assessment conditions matched. It records workload as a secondary outcome and documents when a response is prompted or confirmed by an instructor. The offline tutor tests a referral behavior separately from explanation accuracy. Good referral can preserve a human role while still permitting the software to support comprehension.

> **PROTOCOL APPLICATION**
>
> If effort differs between conditions, record the size and direction of that difference even when the trial also finds improved accuracy.

Validation and source-bound systems

Corso and Koren supply Stanford precedents for designing an explicit validation problem and specifying what a simulation result supports. Their defense records provide high-level methodological context. Asmar’s dissertation record provides a precedent for separating suggestions from the decision maker. None is a direct study of aviation education. [R11, R12, R13]

In the proposed benchmark, the key threat is a fluent answer that lacks support or fails to recognize missing evidence. Freeze versions and capture errors against a pre-agreed rubric. This approach can reveal specific limitations before any learner study. It also allows subsequent teams to rerun a fixed behavioral comparison when a dependency or source revision changes.

> **PROTOCOL APPLICATION**
>
> A traceable source can still have the wrong scope. The stress cases distinguish validity of the citation from validity of the inference.

Learning evidence for AI and XR

Kestin and colleagues tested AI tutoring in a specific undergraduate educational setting. That study motivates an experimentally tractable comparison and offers no estimate for adult pilot training. The XR review by Somerville and colleagues identifies gaps in robust transfer studies and notes concerns such as safety, design quality, and user experience. [R06, R07]

The manuscript therefore separates three questions: can the system produce valid feedback, can learners retain useful reasoning, and does it travel into a different interface? These are distinct claims. A compelling simulator demonstration can be valuable for feasibility, while delayed and independent performance still require controlled measurement.

> **PROTOCOL APPLICATION**
>
> A controlled comparison of tutor feedback should precede a change from desktop to an immersive setting.

Evidence synthesis and gap

The record supports designs that isolate an intervention, structure uncertainty, and measure transfer under changed conditions. It does not establish that a particular commercial tutor improves outcomes for student pilots. The gap is most useful when framed in the language of a decision the research team can take. Is an adaptive feedback feature worth introducing into one lesson if the goal is better independent reasoning two weeks later?

Pilot education also requires enough fidelity for a learner to make the intended inference. The proposed setting prioritizes valid information dependencies rather than visual realism. Further device comparisons are justified only when an independent panel agrees that a spatial task cannot be assessed fairly with the desktop baseline.

> **PROTOCOL APPLICATION**
>
> The proposed causal estimate is for a specific cohort and intervention. The scholarly contribution also includes a way to test its portability.

Aviation Infinity as intervention context

Allem describes Aviation Infinity as a platform that grew from electronic flight tools into AI-supported theory preparation. Its public case study and his *Aviation Futures* articles supply the rationale for evaluating adaptive feedback and instructor handoff. The researcher should obtain versioned test access and document exactly which behaviors the study uses. [R05, R09]

The core trial needs one tightly bounded lesson. Keep objectives, time budget, reference materials, and standard feedback constant, then specify how the AI condition adds adaptation or source-grounded turns. Backend instrumentation can describe delivered intervention, but the estimate concerns random assignment. Any experimental modifications to the product should be labeled as study configurations.

> **PROTOCOL APPLICATION**
>
> The version log should connect lesson pages, prompts, help messages, and materials the typical learner actually encounters.

Avioyx as separate assessment context

Allem’s case study describes Avioyx as a planning and pilot workflow platform. The proposed use is a fictional sandbox with no connection to real flight, booking, or operational actions. Cases contain deliberately inconsistent or stale records and require identifying what needs confirmation. The goal is to test whether the judgment taught in the lesson appears in a different layout. [R10]

If private sandbox access cannot be verified or shared, a neutral open representation of the same assessment can preserve the construct. A small independent scoring study should compare whether both interfaces elicit comparable explanations before formal pooling. The transfer estimate must state the actual interface and version used.

> **PROTOCOL APPLICATION**
>
> If the available planning sandbox differs materially from the proposed configuration, evaluate instrument equivalence before interpreting its scores.

Three supporting aviation cases

AvioSharing illuminates how an interface communicates responsibilities. It supplies a design case on distinguishing required verification from reassuring interface copy. Want To Be a Pilot adds a mentorship context for asking what an instructor needs to know at handoff. New Pilot Shop supports a non-operational procurement case on identifying compatibility evidence. [R16]

These products belong in the background and instrument design. They are not experimental arms. Connecting them to artifact design preserves the intellectual role of the portfolio without claiming that all products must be studied or that product involvement counts as evidence of impact. A new team can reconstruct the tasks from public descriptions and synthetic materials.

> **PROTOCOL APPLICATION**
>
> Supporting case studies supply examples of design alternatives. Their role does not depend on clinical-style performance claims or cross-product comparison.

Open simulator and dynamics dependencies

FlightGear is an open-source simulator. JSBSim is an open-source flight-dynamics library. They are candidate dependencies for later simulator extensions, with separate code and asset notices that must be captured for any redistribution. The primary transfer instrument does not require a real aircraft model and can run as a desktop document-based task. [R17, R18]

This boundary is deliberate. Scoring record interpretation should not depend on graphics performance, network latency, or flight-control skill. If a lab later adds an immersion or dynamic scenario, it should bind the simulator version, dependency versions, assets, and hardware configuration in a manifest and re-establish measurement comparability.

PROTOCOL APPLICATION

Record the exact executable and assets before a simulator is added. The released baseline uses neutral paper-style materials and requires no simulator installation.

Public context data and synthetic fixtures

OurAirports publishes public-domain data and warns about use and accuracy. The project may use such data to understand the structure of airport records, but this release uses only fictional narrative fixtures. Participants can therefore reason about provenance without the materials becoming navigational resources. [R19]

Each assessment pair records both a source conflict and an expected decision explanation. The repository companion provides a machine-readable representation of all cards. An instructor should check whether the evidence presented is sufficient to score it fairly. A later release can add independently reviewed cases, retaining IDs for reproducibility.

> **PROTOCOL APPLICATION**
>
> A fictional register lets raters evaluate the relation between a claim and a specific source without depending on previous knowledge of an airport.

Licensing and rights inventory

A public website, a published dissertation, a software repository, and author-owned product documentation carry different terms. The study can link the source records and distribute original synthetic artifacts. Each release should record the specific terms relevant to any bundled code, dependencies, images, assessment materials, and hardware documentation. [R01, R04, R17-R19]

Permission to use a proprietary platform for testing belongs in a written access agreement. It should cover test accounts, stored prompts, logging, access revocation, and release review, while preserving the investigator’s ability to share defensible findings. Legal and ethics conditions should be evaluated against the actual software and institution involved.

PROTOCOL APPLICATION

A public link to a dissertation or repository is distinct from distributing its PDF, code, or supporting materials.

Versioning and reproducible comparisons

A study record should identify the learning materials, tutor configuration, reference sources, assessment form, scoring rubric, and analysis version at every milestone. A version label alone is inadequate if the related prompt or reference corpus changes silently. Store identifiers or content hashes where feasible and provide a human-readable change log.

Analysis and assessment files can remain portable through stable IDs. The open package supplied here demonstrates checking those IDs and rendering synthetic score summaries. It has no connection to Aviation Infinity or Avioyx accounts. Research staff would create a controlled compatibility layer only after the actual platform access and data permissions are approved.

PROTOCOL APPLICATION

Hash records and configuration versions separately. The resulting linkage makes future differences easier to diagnose.

Founder participation and role separation

The candidate’s product experience is a research asset and a relevant competing interest. A founder can describe what the intervention is meant to do, while independent staff manage recruitment, scoring, and checks on adverse results. The committee should decide who can see unblinded material and what the product team receives. Report the separation in the manuscript and in any related publication. [R05, R20]

Access should be narrow and documented. Subject contact information belongs with recruitment staff. Scorers receive redacted answers and item IDs. Analysts receive arm codes until the analysis outputs are fixed. Product contributors can address technical defects without receiving the comparative outcome tables that might reveal the stronger condition.

> **PROTOCOL APPLICATION**
>
> When a recruiter or rater has a commercial role, identify the role and the mitigation before collecting data.

Define the artifact and test claim

The evaluated artifact is one frozen tutor configuration for one bounded ground-school objective. It includes a model version, system instructions, reference corpus, retrieval settings if used, and a named interface. The test asks whether output meets a predetermined standard of classroom feedback under specified scenarios. It does not evaluate broader certification, operational advice, or every possible prompt.

Separate the product’s interface behavior from the pedagogical content. A helpful display can still expose unsupported or out-of-scope model text. The judging task therefore uses independently extracted, redacted outputs with response format normalized. The operational user interface is separately documented in the intervention record.

> **PROTOCOL APPLICATION**
>
> Use a configuration identity to choose the relevant evidence if someone asks about a later version of the product.

Create and freeze the prompt bank

The offline prompts represent source conflict, uncertainty, missing inputs, inappropriate generalization, and a request for a final decision without enough information. There should also be straightforward answerable items, or a system could appear reliable simply by referring every question away. Hold out cases by stimulus details and passage assignment. Raters review each item for ambiguity before release.

The appendix provides prototype case cards and expected decisions. The final confirmatory benchmark should be independently authored or screened and stored in a restricted repository. Development prompts may guide design. The frozen evaluation prompts must never be used to tune the tutor or to select a model configuration.

> **PROTOCOL APPLICATION**
>
> Conceal benchmark prompts from the team tuning the intervention. Public examples demonstrate the construct for future protocol authors.

Select the comparator

The comparator is a fixed set of instructor-designed feedback and source references for each prompt, prepared using the same objective and source pack. Compare the treatment and comparator within prompt. This paired design controls question difficulty and makes it easier to identify which system output fails on the same underlying issue.

The comparator may also contain errors. Have reviewers score both arms against a separate adjudicated reference standard. Keep presentation consistent where possible while preserving the underlying response. This tests a realistic instructional question and avoids confusing a cross-prompt difference with an advantage of the AI system.

> **PROTOCOL APPLICATION**
>
> The comparator should allow a learner to consult the same citation materials. Control condition quality is part of the research question.

Severity and calibration

Severity categories should depend on whether an error could create a consequential misconception in the educational scenario. An incorrect source title may be minor if the source remains identifiable. Endorsing a conclusion when a required condition is unknown is more serious. Severity judgments should be calibrated with examples before real scoring.

The research team should also distinguish “accurate refusal” from “unnecessary referral.” A tutor that refuses a clear classroom question adds friction and may harm learning. The bank should include answerable items, legitimately incomplete items, and questions where uncertainty can be described even if a final answer is unavailable.

PROTOCOL APPLICATION

Score correctness and appropriate referral separately so a model cannot pass by refusing answerable prompts.

Independent scoring design

Two instructors independently score source support, answer correctness, scope, and referral appropriateness. Record the first pair of ratings before adjudication so agreement is observable. A third reviewer examines disagreements without seeing the system label. Document any item where two plausible interpretations remain after reading the underlying source.

The analysis reports numerator and denominator for each severity category, disagreement patterns, and uncertainty. An overall average can conceal rare serious issues. Therefore the gate uses a predetermined consequential-error criterion and separately reports how frequently the tutor declines answerable questions.

PROTOCOL APPLICATION

Report both initial inter-rater agreement and the final adjudicated conclusions to preserve information on the instrument.

Readiness gate and maintenance

A credible gate names what would stop the study and what would allow repair. Before opening results, the panel chooses a tolerable consequential-error threshold and the precision needed to evaluate it. Requiring an interval bound below the threshold is more informative than treating a small raw error count as proof of reliability.

A failure or uncertain result leads to documented repair and reassessment on a new holdout. The original failed run is part of the research record. A material change to instructions, model, or corpus changes the intervention and requires evidence that the behavior remains suitable for the planned participant setting.

PROTOCOL APPLICATION

When a modified tool fails to reach the intended precision, improve the benchmark and document how the decision rule was chosen.

Stress cases and adverse patterns

The case families include a plausible stale record, mutually inconsistent sources, a silent unit change, an undocumented alternative, and situations where a source has the wrong scope. These are features of records, not real emergency procedures. A response should identify the evidence gap, describe uncertainty, and suggest the next verification within the fictional case.

Review false confidence and false reassurance explicitly. A source-grounded system can still cite a source that fails to support its statement. Preserve enough traces to evaluate that mismatch while protecting any restricted material. The project should also record where the evaluation itself failed to ask a discriminating question.

> **PROTOCOL APPLICATION**
>
> A deliberately ambiguous prompt should invite a qualified explanation, with any forced final answer marked for review.

Outputs and chapter interpretation

Study 1 yields a versioned bank, an expert reference key, paired system outputs, scorer records, and a readiness memo. The public release can include synthetic analogues and the scoring scheme while protecting the held-out trial items. The team must also record item exclusions and why they were made.

The interpretation chapter should explain how the evaluated tasks represent the intended learning objective and where extrapolation would fail. A pass supports the decision to run the controlled lesson under the tested conditions. It provides no direct evidence about learning, operational safety, or performance in a different software version. [R12, R13]

> **PROTOCOL APPLICATION**
>
> The report should include a short table of all consequential error classes and the source of each example.

Population and recruitment

The proposed target population is consenting adult student pilots who have enough ground-school background to interpret the source pack and have not used the frozen study lesson. Recruit across participating training providers only after agreements and institutional review. Describe training stage, previous AI use, and accessibility needs sufficiently to assess representativeness without collecting unnecessary personal details.

Participation must have no effect on a student’s grade, certification, access to training, or relationship with the product. Use staff independent of the founder’s commercial team for invitation and consent. If a class is recruited together, record the grouping and assess whether clustering or interference changes the design.

> **PROTOCOL APPLICATION**
>
> Offer one route to contact the independent coordinator about withdrawal or the use of existing records.

Allocation and concealment

Collect baseline performance before assignment. Within each participating site and predefined experience stratum, use an independently generated allocation sequence with concealment until assignment. Preserve a hash and a dated record of the sequence with the independent coordinator. Maintain a procedure for cancellations, accidental disclosure, and reassignments.

The participant may recognize whether feedback is conversational. Mask scoring by giving reviewers standardized item IDs and plain responses, with identifying or condition-revealing text handled according to a predetermined masking rule. Keep analysts masked to which coded arm received the tutor until the primary output is frozen.

PROTOCOL APPLICATION

Conceal the assignments until baseline scoring is available. Preserve the timing of consent, baseline, and randomization separately.

Keep the intervention contrast legible

Both arms receive the same objectives, classroom source pack, study duration, and opportunities to contact an instructor. The treatment permits a source-grounded tutor to adapt questions and feedback within the specified boundaries. The comparator provides an equivalent instructor-written sequence and static feedback. The research question concerns this intervention package.

Log intervention exposure, downtime, number of replies, and exceptions. These data are useful for fidelity analysis, not grounds to quietly remove participants from the primary comparison. A separate adherence analysis can be exploratory and should state how engagement may confound comparisons between people who complete more or less of the material. [R21]

> **PROTOCOL APPLICATION**
>
> Treatment and comparator materials may differ in feedback mechanics. Instructor availability and source access should remain comparable.

Day-14 endpoint and scoring scale

The proposed primary endpoint is the mean percentage of attainable rubric points on scheduled novel 2D tasks at day 14, with a proposed three-day window in either direction. Each of four criteria receives zero, one, or two points. Before recruitment, calibrate how many tasks are required and define how to handle missing tasks and late visits.

A delayed test should be administered with the same access to the reference pack in each arm, and with no study tutor. Rotate parallel assessment versions after checking equivalence. Score recognition, assumption, traceability, and action separately. This helps diagnose whether improvements reflect better evidence reasoning or only better recognition of familiar item surfaces.

> **PROTOCOL APPLICATION**
>
> The scoring manual must state whether a question can be omitted, how a missing response differs from a zero, and what constitutes a completed visit.

Sample justification and precision

A pilot preceded by the required institutional review should estimate score dispersion, site recruitment, missingness, and rater workload. An independent instructor panel determines the smallest mean difference that would justify adopting the intervention. Use that difference and plausible variance scenarios to justify a planned sample and attrition allowance.

The current proposal deliberately supplies no recruitment target as a finding. Report the design assumptions alongside any chosen target. If multiple recruitment sites create clustering, plan the estimation or allocation accordingly. If the recruitment path is too narrow, revise the design and record the changed precision rather than retrofitting a success threshold after seeing outcomes.

> **PROTOCOL APPLICATION**
>
> Choose the meaningful difference before the investigators have comparative results. Document how educators arrived at it.

Primary analysis and missingness

The primary estimand is the difference in mean delayed rubric percentage by assigned arm in the enrolled eligible population. A prespecified regression can adjust for baseline score and allocation strata. Report a confidence interval, the model specification, and a raw comparison as a transparent check. Preserve a fixed decision about variance estimation and handling of clustering.

Describe missingness by arm and reasons when available. Define the primary missing-data strategy before unmasking. Include a sensitivity analysis that asks how different the missing outcomes would need to be to reverse the substantive conclusion. A complete-case comparison can diagnose sensitivity but should not silently replace the registered primary analysis.

> **PROTOCOL APPLICATION**
>
> Keep raw and adjusted summaries available. Describe what changes when missing scores are modeled under less favorable assumptions.

Secondary outcomes and burdens

Measure immediate performance, study burden, perceived clarity, access barriers, and consequential misconceptions. Workload may change even if accuracy improves. Access indicators can include device interruptions, time to complete an activity, and self-reported difficulty. These indicators should inform implementation and explain exclusions from the studied population.

Report the direction and uncertainty of secondary estimates with an explicit analysis hierarchy. Separate optional inquiry into desktop 3D or headsets as a later comparison. A feasibility report can describe motion discomfort and usability issues without claiming learners improved. Such a comparison should preserve instructional equivalence where the question concerns modality.

> **PROTOCOL APPLICATION**
>
> Secondary outcomes illuminate burden and implementation. They should not be substituted for the confirmatory score.

Flow, deviations, and reporting

Report how many were approached, assessed, randomized, exposed to the assigned materials, followed up, and analyzed. The ledger should account for timing deviations, differential interruptions, and fixes made after the freeze. Keep denominators visible. Report the raw and adjusted primary estimates and an interval, whatever their direction. [R20, R21]

Jaedicke's Chapter IV reports sample characteristics, pilot observations, assumptions, and measured outcomes separately. Use that reporting sequence here so readers can distinguish changes in the sample from changes in performance. A small estimate can inform instructors even if precision is high. A wide interval spanning both reasonable benefit and harm warrants different language. The manuscript should explain each case before outcome analysis. An open draft table is a useful contract because it shows the numbers readers should be able to see later.

> **PROTOCOL APPLICATION**
>
> The flow ledger should separate participants who cannot complete the delivery format from those who decline follow-up.

Transfer design and independence

Study 3 uses the same randomized participants and a separate, held-out simulated planning interface. Its endpoint is secondary to the day-14 lesson assessment. The transfer task gives both arms the same static references, the same time window, and no access to the study tutor. Assign task order independently and preserve the order in the data.

Assessment features must change enough to test learning across contexts while preserving the construct. Change the presentation layout, entity names, wording, and evidence location. Retain the need to reason from a dated source and explain the relevant uncertainty. The independent assessment may run through Avioyx or a documented neutral sandbox. [R08, R10]

> **PROTOCOL APPLICATION**
>
> The assessment starts on a clean sandbox and uses identical reference access for both arms.

Scenario families and leakage prevention

Use simulated record conflicts covering update dates, source scope, labels, missing approvals, and assumptions. The public case cards in this draft illustrate candidate families. Actual trial forms require a separate protected bank. Keep field items away from lesson authors and the model evaluation team to avoid accidental exposure.

Record a lineage identifier and fingerprint for both lesson and assessment materials. A reviewer should identify items that differ only in superficial details. Decide in advance whether an overlap invalidates a single item or the corresponding form. Keep the treatment and control conditions equally insulated from the held-out bank.

> **PROTOCOL APPLICATION**
>
> Review scenario overlap independently through evidence structures as well as duplicate wording.

Score reasoning in the planning context

An assessor reads the written answer against four criteria. The first criterion asks whether the participant identifies the relevant discrepancy. The second asks whether the explanation states its decision significance. The third tracks the source and the fourth requires a defensible verification step. Score the response a learner actually supplied, even if it differs from the example wording.

Midpoint examples help raters distinguish incomplete and contradictory answers. Allow alternate sources or actions if the evidence makes them defensible. Require a procedure to resolve genuinely ambiguous prompts. Exclude such an item consistently from both arms and record the reason independently of how much the score changes.

> **PROTOCOL APPLICATION**
>
> When a conditional response is well supported, allow it even if it differs from the sample response.

Configuration and pilot validation

Use fictional flights, sources, values, and identifiers. Freeze the software version and record language, display, browser, device configuration, and scripted instructions. Confirm that failures of the sandbox do not resemble participant errors and define how to log an interrupted task.

Ask an independent instructor panel to review whether each form elicits the intended construct. Conduct a dry run using non-study responses and request a structured review of usability and form equivalence. If comparable operation across providers cannot be established, state the limitation. Configuration testing is feasibility evidence and precedes effectiveness claims.

> **PROTOCOL APPLICATION**
>
> Give reviewers a consistent assumption about whether notes, sources, and the scoring timer are available.

Workload, confidence, and disagreement

Record response time and self-reported workload only after protecting the primary assessment. Timing can reflect familiarity with a tool even when the participant reasons correctly. Confidence can remain high after a partially wrong answer, or be appropriately low when records conflict. Analyze these relationships with care.

Report score agreement before adjudication and track where one dimension shows more rater disagreement. This can indicate that the task or rubric needs refinement. The dissertation can include a bounded subset of cleared examples if permitted, while the repository can offer fictional replacements to demonstrate the same scoring principles.

> **PROTOCOL APPLICATION**
>
> A faster answer can be weaker. Preserve time and reasoning score as separate fields.

Secondary analysis and dependence

Since participants are the same in both main assessments, the transfer comparison is not an independent second randomized trial. Preserve the same assignment and show which participants returned for each assessment. Define the summary score and the method for baseline adjustment, then maintain separate tables for Study 2 and Study 3.

Report subgroup patterns as exploratory unless specifically powered and registered. Compare attrition and operational interruptions by arm. Explain uncertainty and avoid overstating replication because a second interface still relies on the same participant pool. A later cohort or independent team can test replication directly.

PROTOCOL APPLICATION

Use a prespecified interpretation plan for correlations between the two assessments. They share randomized participants.

Alternative explanations and falsification

A favorable score on the 2D follow-up with no improvement in planning can indicate that the learned skill is too narrow, the planning instrument is noisy, or the delay and task mix differ. Test these possibilities using recorded instrument characteristics and blinded qualitative summaries. A similar pattern across both screens can still reflect the particular materials selected.

Negative controls can include straightforward source lookup items where the intervention has little expected advantage, provided they do not affect the primary estimand. Define those checks in advance. If results are uniformly favorable on every measure, investigate scorer unmasking or task exposure before interpreting them as learning.

PROTOCOL APPLICATION

An observed transfer effect still needs scrutiny for familiarity, scorer unmasking, attrition, and inequivalent forms.

Evidence package for transfer

The planned release includes the final instrument version permitted for sharing, a score dictionary, verified aggregate comparisons, and a synthetic scenario bank to demonstrate use. Preserve assessment items with restricted reuse separately from public artifacts. Document which simulator components were actually used.

A replicator should be able to identify the study configuration and build an equivalent sandbox given the version record. If that is impossible, explain the dependency and provide a neutral synthetic demonstration. Keep product release claims separate from conclusions drawn under the study’s research and simulation conditions.

PROTOCOL APPLICATION

Replicators should be able to read how the tool, task bank, reference material, and scoring manual were linked for a given result.

Ethics and consent pathway

Before recruiting or using participant records, obtain the appropriate written institutional determination and site agreements. The protocol should explain assignment, what data are collected, possible inconvenience, withdrawal procedure, and the research team’s ties to the products. A participant should be able to decline without affecting training.

Capture only the variables needed to answer the question and to describe the scope of participation. Document permissions for data handling and dissemination. Separate participant identifiers from the analysis data. Unexpected adverse effects and privacy events require a documented reporting route and a defined decision about pausing study activity.

PROTOCOL APPLICATION

The study should maintain an independent contact and an approved process for concerns, incidents, and withdrawal.

Accessibility is an outcome and a design condition

Use a desktop-first interface with keyboard navigation, readable contrast, non-color-dependent labels, and suitable textual alternatives for visual material. Give the participant reasonable opportunity to use approved accommodations. Record which design changes preserve the construct and which require a separate feasibility check.

A headset comparison would need an accessible alternative and a clear spatial rationale. Record discomfort, access barriers, and who was excluded by the device requirements. Failure to participate in an interface is part of its implementation burden. It should not be disguised as a missing entry in a table of successful completions.

PROTOCOL APPLICATION

A record of nonparticipation can clarify whether an intervention is feasible for people the team intended to reach.

Distinguish research assistance from intervention

AI may assist with literature discovery, outlining and drafting, code, analysis development, and reproducibility checks. The study intervention is a specific tutor configuration presented to selected participants. These activities have different version and data records. Logging them separately prevents the dissertation from confusing assistance in writing with the object of measurement. [R22]

The dissertation author verifies assertions, sources, computations, and interpretations. Substantial AI assistance should be disclosed in the submitted document. If tools receive restricted material, approved environments and documented permissions are required. Human raters and analysts retain their defined responsibilities and independently review material outputs.

PROTOCOL APPLICATION

Log AI assistance in investigation separately from the tutor version and setting presented to trial participants.

Source verification and citations

A tool can suggest a plausible citation without validating it. Check bibliographic identity and directly inspect the claim-bearing source. Distinguish access to an abstract from access to a full dissertation. Tie claims to the most direct source actually reviewed.

The source register demonstrates this difference by labeling dissertation catalog records, defense materials, research articles, and product descriptions. Update it before submission if additional access is obtained. The reference list should retain the selected edition and a stable link with an access date where useful.

PROTOCOL APPLICATION

Use public bibliographic facts to support the cited claim, and verify any proposed quotation against its original source.

Data minimization and retention

Restrict research data to consented fields and establish a written retention schedule. Merge de-identified files using a pseudonymous participant ID. Preserve the least intrusive log needed to document exposure and verify analysis. Avoid transmitting identifiable or commercially restricted content into a general-purpose AI tool.

Use access records to show who saw real data and when. Public examples should be generated independently from protected responses. If a dataset cannot be publicly shared, describe the access pathway or a justified restriction and supply a synthetic fixture that demonstrates the structure and analysis.

> **PROTOCOL APPLICATION**
>
> Use synthetic fixtures for public code checks. Keep the participant linking key in a separate controlled system.

Versioned AI disclosure

A material-use entry records date, tool and version where known, purpose, artifact, category of data, and independent verification. If a model version is unavailable, record the limitation and the observed environment. Keep prompts or a sufficient description of the interaction where doing so is permitted. Verify coding suggestions with tests and independently re-run critical outputs.

The candidate prepares a concise disclosure using actual use. This working manuscript was developed with AI assistance. A dissertation receiving review should explain which material the candidate revised, verified, or authored, with corresponding evidence. An illustrative disclosure is supplied in Appendix B.

> **PROTOCOL APPLICATION**
>
> The dissertation disclosure should explain how the author checked AI-generated output and which artifacts were affected.

Study risk versus simulated task content

All assessment tasks are fictional and bounded to educational judgment. The apparatus must prevent consequence-bearing flight actions and use no live bookings or aircraft connection. The study question concerns conditional reasoning and traceability in a simulated context. Ask instructors to explain these boundaries during orientation.

Consent, workload monitoring, and a stopping procedure are part of routine study operations. If a task evokes discomfort or concern, the participant can pause or leave under the approved procedure. Document suspension rules symmetrically across study arms so operational convenience does not drive the comparison.

> **PROTOCOL APPLICATION**
>
> A simulation can test a particular reasoning task with more control than a live environment but supports a narrower claim.

Governance for adverse results

A team can agree before starting that serious errors are reported and that null results enter the scholarly record. Independent sign-off protects this agreement. Prepare a release pathway for restricted material and a clear process for documenting what is withheld and why.

The results chapter should be guided by estimates and intervals. A product-facing brief may describe design priorities, but should align with the dissertation’s limitations. Preserve the dissertation’s right to report negative results within approved confidentiality constraints. The resulting research is more useful when it can correct assumptions about the product.

> **PROTOCOL APPLICATION**
>
> An analysis lock and independent reviewer make it easier to address uncomfortable or disappointing findings openly.

Integrating the three evidence claims

The three studies answer a question in stages. Reliability is a property of the frozen tutor under specified benchmarks. Delayed learning estimates how participants respond to assignment under the trial conditions. Transfer checks whether that difference appears in a separate interface without the study tutor. Discuss how each supports or limits the next.

The synthesis should make this chain visible through task, model, source, and rubric versions. A positive comparison matters most when an outside reader can see who participated, which materials they received, and what independent assessment tested. Claims should follow the narrowest relevant evidence.

PROTOCOL APPLICATION

Align each potential claim with the exact study, setting and version that supports it.

Interpreting positive and equivocal findings

If the estimated effect meets the prespecified educational threshold with useful precision, discuss the conditions under which that result holds. Evidence of a difference at day 14 supports a delay-specific claim. It does not automatically establish persistence at later times or widespread effectiveness across unrelated objectives.

If the interval contains both trivial and worthwhile effects, distinguish inconclusive evidence from evidence of a small effect. If results favor the comparator, investigate the nature of the tutor’s guidance and workload before proposing redesign. Report what the data actually support even when the preferred commercial outcome is absent.

> **PROTOCOL APPLICATION**
>
> An equivalence claim requires an explicit design and justified margin. A non-significant difference alone is insufficient.

Interpretations when stages diverge

A reliable tutor and no measurable learning benefit imply different next steps from an unreliable tutor plus a promising early practice result. In the first case, consider whether the helpful response remains too similar to static feedback. In the second, address the benchmark defect before exposing additional learners. Separate design failure, fidelity failure, and uncertain estimation.

If only the planning task shows a difference, examine the prespecified analysis, scoring comparability, and task composition. A single secondary finding is hypothesis-generating unless the study plan provides a basis for stronger inference. Replication in a different cohort can test whether the pattern travels.

> **PROTOCOL APPLICATION**
>
> Use a prespecified hierarchy: reliability, delayed learning and separate simulated transfer.

Generalizability and external validity

The planned recruitment and simulation choices set the scope of inference. A result from adult student pilots using one curriculum and one source pack cannot be directly transferred to professional pilots, younger learners, another licensing setting, or live operation. Report participation, setting, and technology clearly enough for another team to make that judgment.

A follow-up study could vary instructional provider, learner experience, or source pack while preserving the outcome definitions. Compare task versions and teach the same learning objective. A revised trial for a spatial skill could deliberately vary modality. Changes to both feedback and modality at once should be interpreted as a package comparison.

> **PROTOCOL APPLICATION**
>
> A setting can be representative of a training objective without being representative of all pilots or real flight conditions.

How the open benchmark can be reused

The supplied scenario bank offers synthetic demonstrations of the scoring construct, with instructor-facing notes and case IDs. The accompanying Python script validates IDs and scores, then reports only a summary for its built-in synthetic examples. It illustrates the data shape and is not the trial estimator.

An independent team can run the package without access to either proprietary product. Extend it by keeping a public development bank and a protected assessment bank separately versioned. Record changed task features and grading consequences. Use clinical-sounding or operational language only when it accurately reflects the simulator’s research setting.

> **PROTOCOL APPLICATION**
>
> The companion fixture tool checks input assumptions and scoring mechanics. It does not implement a final clinical-style trial analysis.

Dissertation defense

A defense should show the central question, why the design isolates a meaningful comparison, and where the outcomes stop supporting further claims. Demonstrate the independent scoring flow and one synthetic reproduction. Explain one non-obvious design choice, such as fixing the interface before testing whether a modality change is useful.

The candidate should be ready to defend the numerator and denominator for the primary outcome, the missing-data plan, and the role of commercial interests. Bring the actual AI disclosure and version records. If the proposed work has not been completed, the relevant event is a proposal defense, with each pending requirement clearly identified.

> **PROTOCOL APPLICATION**
>
> In the defense, use one complete trace from a research question through the source, measure, result and limitation.

Publication and deposit plan

First archive the approved protocol and any shareable code, then date outcome tables derived from the locked analysis. Write manuscripts around defensible contributions and follow the editorial policies of the receiving journal. The proposal and previously published perspectives can frame a program of research without standing in for trial results.

Prepare an integrative dissertation with methods, results when available, limitations, documentation of candidate contributions, and the reproducibility statement. Keep repository release terms, restricted-data contacts, and instructions for independent review together. A final deposited version follows through the designated institutional channel after approval.

PROTOCOL APPLICATION

Deposited materials require approved clearance and version records for any code, data, or assets shared.

Conclusion

The question is practical for an instructor: after the tutor leaves, what can the learner still explain? The dissertation design pairs a study of reliability with a short controlled lesson and a delayed check in a second interface. It gives independent scorers a defined construct and gives another team materials to test whether the method can be reproduced.

The design does not require a positive result to be valuable. A careful negative or mixed result would identify where AI feedback contributes little, where it changes behavior, and what should be tested next. The strongest final contribution will be the candidate’s defensible interpretation of observed evidence, supported by a reproducible record.

> **PROTOCOL APPLICATION**
>
> Include interpretation of negative, adverse, and uncertain outcomes in the eventual discussion.

Training design and transfer

R01

Jaedicke, C. K. (2026). Enhancing Virtual Reality Cockpit Procedure Training with Haptic Feedback. Embry-Riddle Aeronautical University doctoral dissertation.

https://commons.erau.edu/edt/1016/

Full-text supplied PDF and public catalog. Consulted the abstract, Chapter III methods structure, Chapter IV reporting structure, Chapter V limitations, and appendix register.

R02

Zhang, T. (2022). Comparing Training Effects of Virtual Reality Flight Simulation to Conventional PC-Based Flight Simulation. Embry-Riddle doctoral dissertation.

https://commons.erau.edu/edt/668/

Repository abstract motivates an assessment environment distinct from the instructional treatment.

R03

Fussell, S. G. (2020). Determinants of Aviation Students' Intentions to Use Virtual Reality for Flight Training. Embry-Riddle doctoral dissertation.

https://commons.erau.edu/edt/542/

Repository abstract supplies an adoption precedent. Keep intention to use distinct from observed learning.

Automation and educational effectiveness

R04

Johnson, A. W. (2015). Analyzing the effects of dynamic task allocation on human-automation system performance. MIT doctoral dissertation.

https://dspace.mit.edu/entities/publication/d306899d-0509-43e6-a4cc-cceb09...

Repository record and abstract support separating workload, attention, and performance.

R05

Allem, A. (2026). Subject-Specific AI Instruction Across 2D, Desktop 3D, and Headset-Based Aviation Learning. Aviation Futures, 2(1).

https://doi.org/10.68956/journalos.jrn-2026-ai-a965b42873984c028ee7e2f7270...

Published conceptual framework proposes evaluating objectives, tutoring, delivery modes, and human oversight. See the journal’s author record for related work.

R06

Kestin, G., et al. (2025). AI tutoring outperforms in-class active learning: an RCT introducing a novel research-based design in an authentic educational setting. Scientific Reports.

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

Relevant design precedent in undergraduate education. It does not supply an effect estimate for pilot training.

Transfer, tutoring, and platforms

R07

Somerville, A. D., et al. (2025). Applications of extended reality in pilot flight simulator training: a systematic review with meta-analysis.

https://doi.org/10.1186/s42492-025-00206-w

Research synthesis identifies limits in robust transfer evidence and justifies a bounded simulator assessment.

R08

Barnett, S. M., and Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. Psychological Bulletin, 128(4), 612-637.

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

Taxonomy informs specification of context, presentation, knowledge domain, and time delay.

R09

Allem, A. Aviation Infinity. Public product case study.

https://kingallem.com/case-study/aviation-infinity

Author-described materials and technology context. The trial must verify the actual delivered configuration independently.

Portfolio and Stanford precedents

R10

Allem, A. Avioyx. Public product case study.

https://kingallem.com/case-study/avioyx

Author-described planning interfaces. The candidate study uses only fictional sandbox cases.

R11

Asmar, D. M. (2025). Collaborative Multiagent Decision Making via Action Suggestions. Stanford University doctoral dissertation bibliographic record.

https://dylanasmar.com/publications/

Author-hosted record. Provides a precedent for considering suggestions separately from decision responsibility.

R12

Corso, A. (2021). Algorithms for Black-Box Safety Validation. Stanford dissertation defense slides.

https://web.stanford.edu/group/sisl/public/defense_corso.pdf

Defense material for a Stanford dissertation, used for validation design context.

Stanford validation and aviation context

R13

Koren, M. (2021). Approximate Methods for Validating Autonomous Systems in Simulation. Stanford dissertation defense slides.

https://web.stanford.edu/group/sisl/public/defense_koren.pdf

Defense material for a Stanford dissertation, used for simulation validation as a methodological precedent.

R14

Federal Aviation Administration. Aviation Instructor’s Handbook.

https://www.faa.gov/regulations_policies/handbooks_manuals/aviation/aviati…

Aviation education reference and a starting point for a source pack. Instructors should confirm selections and editions.

R15

Schlichting, M. R., et al. (2025). LeRAAT: LLM-Enabled Real-Time Aviation Advisory Tool. arXiv:2503.16477.

https://arxiv.org/abs/2503.16477

Related Stanford-affiliated research on a simulator advisory system. The proposed transfer assessment here removes access to the tutor.

Design applications and tools

R16

Allem, A. Projects by Ahmed Allem. Public portfolio.

https://kingallem.com/case-studies

Public design context for the aviation, mentorship, and product comparison cases.

R17

FlightGear Project. FlightGear software and license.

https://www.flightgear.org/about/license/

Open-source candidate simulator. Record licenses of selected code and any individual bundled assets before redistribution.

Open tools and evidence reporting

R18

JSBSim Team. JSBSim flight dynamics library.

https://github.com/JSBSim-Team/jsbsim

Open-source candidate library; review pinned source and asset terms before integration.

R19

OurAirports. About OurAirports and available data.

https://ourairports.com/about.html

Public-domain data are a possible example of schema structure; the included scenarios use synthetic records.

R20

Hopewell, S., et al. (2025). CONSORT 2025 statement: updated guideline for reporting randomised trials. BMJ, 389, e081123.

https://www.bmj.com/content/389/bmj-2024-081123

Supports transparent participant flow, allocation and deviations, and complete outcome reporting.

Reporting, research integrity, and catalogs

R21

Hoffmann, T. C., et al. (2014). Better reporting of interventions: template for intervention description and replication. BMJ, 348, g1687.

https://www.bmj.com/content/348/bmj.g1687

Checklist context for documenting the intervention and comparator in reproducible detail.

R22

University of California, Santa Barbara Graduate Division. Guidance on responsible and ethical use of AI in graduate academics.

https://www.graddiv.ucsb.edu/resources/guidance-ai-graduate-academics

External guidance on verification, disclosure, and researcher responsibility. Research.University determines its own approvals and examination expectations.

R23

Stanford University Libraries. Dissertation and thesis submission.

https://library.stanford.edu/dissertation-and-thesis-submission-phd-jsd-dm...

Explains discovery through SearchWorks and the Stanford Digital Repository, including records that can be subject to embargo.

Discovery and authorship records

R24

Stanford Intelligent Systems Laboratory. Doctoral graduates and dissertation titles.

https://sisl.stanford.edu/doctoral-graduates/

Official lab index for the Stanford precedents named in the literature review.

R25

Aviation Futures. Ahmed Allem author record.

https://journal.research.university/journals/aviation-futures/authors/ahme…

Public register for the three related publications used as background.

R26

Allem, A. (2026). Aviation Infinity: Learning That Travels to the Cockpit. Aviation Futures.

https://journal.research.university/journals/aviation-futures/articles/avi…

Perspective framing learning continuity and instructor handoff.

Related portfolio publications

R27

Allem, A. (2026). An Evidence Plan for XR Flight Training: From an Inspiring Demonstration to a Defensible Academy Decision. Aviation Futures.

https://journal.research.university/journals/aviation-futures/articles/xr-...

Research adaptation informing the separation between feasibility and effectiveness.

R28

Research.University. Aerospace & Aeronautics PhD: Dissertation dossier.

https://research-university-aerospace-phd.freedom730076.chatgpt.site/disse...

Program context, module links and the earlier proposal. The present manuscript extends the same plan.

A01-01 · Weather bulletin provenance

CURRENT-VERSUS-ARCHIVED

The exercise header shows a planned departure window. A short briefing and an update notice appear in different panels; the summary does not name its source.

CASE STIMULUS

Panel A is marked current in the interface. The register identifies A as an archived copy and B as the applicable revision. Both are accessible.

A: Training bulletin / Harbor arrival window, rev 2, archived. B: rev 3. Register: rev 3 applies to this exercise.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied briefing receipt to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Explain which copy is applicable and cite the register entry and the revision label. Compare the header revision and the date in the source register before accepting the short summary.

> **SCORING BOUNDARY**
>
> Document date and applicability matter more than whether a familiar icon is present. Present both labels in equivalent positions. Mark a response incomplete if it selects B without identifying the evidence or if it confuses recency with applicability. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A01-02 · Weather bulletin provenance

MISSING SCOPE

The exercise header shows a planned departure window. A short briefing and an update notice appear in different panels; the summary does not name its source.

CASE STIMULUS

The first source covers Object A only. The item asks about Object B, and the documentation for B is deliberately absent.

A: Training bulletin / Harbor arrival window, training venue=Alpha. Task: training venue=Beta. Register: no Beta record supplied.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied briefing receipt to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Identify the scope mismatch, describe the missing evidence, and ask for a source explicitly covering B. Compare the header revision and the date in the source register before accepting the short summary.

> **SCORING BOUNDARY**
>
> Document date and applicability matter more than whether a familiar icon is present. A confident generalization to B is a consequential misconception in this exercise. Accept clear conditional explanations. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A01-03 · Weather bulletin provenance

DUPLICATED SUMMARY

The exercise header shows a planned departure window. A short briefing and an update notice appear in different panels; the summary does not name its source.

CASE STIMULUS

Two interface summaries repeat the same claim. They resolve to a single upstream record released under an earlier scenario assumption.

Tile A and tile B cite Training bulletin / Harbor arrival window, origin Q4. Q4 assumes practice session=Alpha. Task fixes practice session=Beta.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied briefing receipt to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Identify the common upstream source. Explain why repetition does not independently establish the claim under the current assumption. Compare the header revision and the date in the source register before accepting the short summary.

> **SCORING BOUNDARY**
>
> Document date and applicability matter more than whether a familiar icon is present. Require evidence that the two summaries share a dependency. Do not award full traceability for naming two tiles. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A01-04 · Weather bulletin provenance

USABLE EVIDENCE

The exercise header shows a planned departure window. A short briefing and an update notice appear in different panels; the summary does not name its source.

CASE STIMULUS

The authoritative source explicitly covers the scenario, is within its validity window, and supports a single bounded conclusion.

A: Training bulletin / Harbor arrival window, applies to this case. Register: A is valid for today's fictional session. A decision field: arrival briefing confirmed.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied briefing receipt to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

State the limited conclusion with its source and explain why the evidence is sufficient for the question asked. Compare the header revision and the date in the source register before accepting the short summary.

> **SCORING BOUNDARY**
>
> Document date and applicability matter more than whether a familiar icon is present. This is an answerable case. Escalation alone is insufficient. Reward agreement limited to the question actually posed. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A01-05 · Weather bulletin provenance

CONFLICTING RECORDS

The exercise header shows a planned departure window. A short briefing and an update notice appear in different panels; the summary does not name its source.

CASE STIMULUS

Two dated records make different assertions and the scenario pack provides no precedence rule. A separate note states which roles can resolve disagreements.

A: Training bulletin / Harbor arrival window, arrival briefing confirmed. B: arrival briefing pending. No priority field. Role register: ask the independent scenario coordinator.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied briefing receipt to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Describe the conflict and identify a verifiable next action without selecting a winner on unsupported grounds. Compare the header revision and the date in the source register before accepting the short summary.

> **SCORING BOUNDARY**
>
> Document date and applicability matter more than whether a familiar icon is present. A response can give conditional implications for each record. It must avoid a final unsupported choice. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A01-06 · Weather bulletin provenance

ASSUMPTION CHANGED

The exercise header shows a planned departure window. A short briefing and an update notice appear in different panels; the summary does not name its source.

CASE STIMULUS

The apparently useful record was generated under a named baseline assumption. The current scenario explicitly changes that assumption.

Record A: Training bulletin / Harbor arrival window, practice session=Alpha. Exercise overlay: practice session=Beta. Revision for Beta is missing.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied briefing receipt to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Name the changed assumption and identify what must be rechecked before reusing the record. Compare the header revision and the date in the source register before accepting the short summary.

> **SCORING BOUNDARY**
>
> Document date and applicability matter more than whether a familiar icon is present. Require a causal explanation of why the original record may no longer apply. Accept more than one reasonable way to request a new check. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A02-01 · Runway-status notice scope

CURRENT-VERSUS-ARCHIVED

A fictional airport bulletin concerns one labeled surface. A dashboard tile uses a general availability label. Both appear in the same scenario pack.

CASE STIMULUS

Panel A is marked current in the interface. The register identifies A as an archived copy and B as the applicable revision. Both are accessible.

A: Surface availability / Mosaic surface label, rev 2, archived. B: rev 3. Register: rev 3 applies to this exercise.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied surface notice to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Explain which copy is applicable and cite the register entry and the revision label. Identify which labeled surface the source actually describes and what is unknown about the other surface.

> **SCORING BOUNDARY**
>
> Scope is a property of a specific record. Do not infer that a notice about one surface applies to every surface. Present both labels in equivalent positions. Mark a response incomplete if it selects B without identifying the evidence or if it confuses recency with applicability. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A02-02 · Runway-status notice scope

MISSING SCOPE

A fictional airport bulletin concerns one labeled surface. A dashboard tile uses a general availability label. Both appear in the same scenario pack.

CASE STIMULUS

The first source covers Object A only. The item asks about Object B, and the documentation for B is deliberately absent.

A: Surface availability / Mosaic surface label, surface scope=Alpha. Task: surface scope=Beta. Register: no Beta record supplied.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied surface notice to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Identify the scope mismatch, describe the missing evidence, and ask for a source explicitly covering B. Identify which labeled surface the source actually describes and what is unknown about the other surface.

> **SCORING BOUNDARY**
>
> Scope is a property of a specific record. Do not infer that a notice about one surface applies to every surface. A confident generalization to B is a consequential misconception in this exercise. Accept clear conditional explanations. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A02-03 · Runway-status notice scope

DUPLICATED SUMMARY

A fictional airport bulletin concerns one labeled surface. A dashboard tile uses a general availability label. Both appear in the same scenario pack.

CASE STIMULUS

Two interface summaries repeat the same claim. They resolve to a single upstream record released under an earlier scenario assumption.

Tile A and tile B cite Surface availability / Mosaic surface label, origin Q4. Q4 assumes simulated activity=Alpha. Task fixes simulated activity=Beta.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied surface notice to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Identify the common upstream source. Explain why repetition does not independently establish the claim under the current assumption. Identify which labeled surface the source actually describes and what is unknown about the other surface.

> **SCORING BOUNDARY**
>
> Scope is a property of a specific record. Do not infer that a notice about one surface applies to every surface. Require evidence that the two summaries share a dependency. Do not award full traceability for naming two tiles. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A02-04 · Runway-status notice scope

USABLE EVIDENCE

A fictional airport bulletin concerns one labeled surface. A dashboard tile uses a general availability label. Both appear in the same scenario pack.

CASE STIMULUS

The authoritative source explicitly covers the scenario, is within its validity window, and supports a single bounded conclusion.

A: Surface availability / Mosaic surface label, applies to this case. Register: A is valid for today's fictional session. A decision field: surface status confirmed.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied surface notice to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

State the limited conclusion with its source and explain why the evidence is sufficient for the question asked. Identify which labeled surface the source actually describes and what is unknown about the other surface.

> **SCORING BOUNDARY**
>
> Scope is a property of a specific record. Do not infer that a notice about one surface applies to every surface. This is an answerable case. Escalation alone is insufficient. Reward agreement limited to the question actually posed. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A02-05 · Runway-status notice scope

CONFLICTING RECORDS

A fictional airport bulletin concerns one labeled surface. A dashboard tile uses a general availability label. Both appear in the same scenario pack.

CASE STIMULUS

Two dated records make different assertions and the scenario pack provides no precedence rule. A separate note states which roles can resolve disagreements.

A: Surface availability / Mosaic surface label, surface status confirmed. B: surface status pending. No priority field. Role register: ask the independent scenario coordinator.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied surface notice to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Describe the conflict and identify a verifiable next action without selecting a winner on unsupported grounds. Identify which labeled surface the source actually describes and what is unknown about the other surface.

> **SCORING BOUNDARY**
>
> Scope is a property of a specific record. Do not infer that a notice about one surface applies to every surface. A response can give conditional implications for each record. It must avoid a final unsupported choice. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A02-06 · Runway-status notice scope

ASSUMPTION CHANGED

A fictional airport bulletin concerns one labeled surface. A dashboard tile uses a general availability label. Both appear in the same scenario pack.

CASE STIMULUS

The apparently useful record was generated under a named baseline assumption. The current scenario explicitly changes that assumption.

Record A: Surface availability / Mosaic surface label, simulated activity=Alpha. Exercise overlay: simulated activity=Beta. Revision for Beta is missing.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied surface notice to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Name the changed assumption and identify what must be rechecked before reusing the record. Identify which labeled surface the source actually describes and what is unknown about the other surface.

> **SCORING BOUNDARY**
>
> Scope is a property of a specific record. Do not infer that a notice about one surface applies to every surface. Require a causal explanation of why the original record may no longer apply. Accept more than one reasonable way to request a new check. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A03-01 · Chart edition comparison

CURRENT-VERSUS-ARCHIVED

A route worksheet and a chart extract share a title. Their edition labels, legend entries, and archived identifiers disagree in a way visible in the provided pack.

CASE STIMULUS

Panel A is marked current in the interface. The register identifies A as an archived copy and B as the applicable revision. Both are accessible.

A: Chart edition / Alder worksheet, rev 2, archived. B: rev 3. Register: rev 3 applies to this exercise.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied chart register to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Explain which copy is applicable and cite the register entry and the revision label. Use the edition register and printed legend to explain the inconsistency; request an applicable extract if necessary.

> **SCORING BOUNDARY**
>
> The task measures evidence versioning. Geographical familiarity should provide no scoring advantage. Present both labels in equivalent positions. Mark a response incomplete if it selects B without identifying the evidence or if it confuses recency with applicability. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A03-02 · Chart edition comparison

MISSING SCOPE

A route worksheet and a chart extract share a title. Their edition labels, legend entries, and archived identifiers disagree in a way visible in the provided pack.

CASE STIMULUS

The first source covers Object A only. The item asks about Object B, and the documentation for B is deliberately absent.

A: Chart edition / Alder worksheet, worksheet edition=Alpha. Task: worksheet edition=Beta. Register: no Beta record supplied.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied chart register to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Identify the scope mismatch, describe the missing evidence, and ask for a source explicitly covering B. Use the edition register and printed legend to explain the inconsistency; request an applicable extract if necessary.

> **SCORING BOUNDARY**
>
> The task measures evidence versioning. Geographical familiarity should provide no scoring advantage. A confident generalization to B is a consequential misconception in this exercise. Accept clear conditional explanations. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A03-03 · Chart edition comparison

DUPLICATED SUMMARY

A route worksheet and a chart extract share a title. Their edition labels, legend entries, and archived identifiers disagree in a way visible in the provided pack.

CASE STIMULUS

Two interface summaries repeat the same claim. They resolve to a single upstream record released under an earlier scenario assumption.

Tile A and tile B cite Chart edition / Alder worksheet, origin Q4. Q4 assumes legend set=Alpha. Task fixes legend set=Beta.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied chart register to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Identify the common upstream source. Explain why repetition does not independently establish the claim under the current assumption. Use the edition register and printed legend to explain the inconsistency; request an applicable extract if necessary.

> **SCORING BOUNDARY**
>
> The task measures evidence versioning. Geographical familiarity should provide no scoring advantage. Require evidence that the two summaries share a dependency. Do not award full traceability for naming two tiles. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A03-04 · Chart edition comparison

USABLE EVIDENCE

A route worksheet and a chart extract share a title. Their edition labels, legend entries, and archived identifiers disagree in a way visible in the provided pack.

CASE STIMULUS

The authoritative source explicitly covers the scenario, is within its validity window, and supports a single bounded conclusion.

A: Chart edition / Alder worksheet, applies to this case. Register: A is valid for today's fictional session. A decision field: chart alignment confirmed.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied chart register to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

State the limited conclusion with its source and explain why the evidence is sufficient for the question asked. Use the edition register and printed legend to explain the inconsistency; request an applicable extract if necessary.

> **SCORING BOUNDARY**
>
> The task measures evidence versioning. Geographical familiarity should provide no scoring advantage. This is an answerable case. Escalation alone is insufficient. Reward agreement limited to the question actually posed. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A03-05 · Chart edition comparison

CONFLICTING RECORDS

A route worksheet and a chart extract share a title. Their edition labels, legend entries, and archived identifiers disagree in a way visible in the provided pack.

CASE STIMULUS

Two dated records make different assertions and the scenario pack provides no precedence rule. A separate note states which roles can resolve disagreements.

A: Chart edition / Alder worksheet, chart alignment confirmed. B: chart alignment pending. No priority field. Role register: ask the independent scenario coordinator.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied chart register to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Describe the conflict and identify a verifiable next action without selecting a winner on unsupported grounds. Use the edition register and printed legend to explain the inconsistency; request an applicable extract if necessary.

> **SCORING BOUNDARY**
>
> The task measures evidence versioning. Geographical familiarity should provide no scoring advantage. A response can give conditional implications for each record. It must avoid a final unsupported choice. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A03-06 · Chart edition comparison

ASSUMPTION CHANGED

A route worksheet and a chart extract share a title. Their edition labels, legend entries, and archived identifiers disagree in a way visible in the provided pack.

CASE STIMULUS

The apparently useful record was generated under a named baseline assumption. The current scenario explicitly changes that assumption.

Record A: Chart edition / Alder worksheet, legend set=Alpha. Exercise overlay: legend set=Beta. Revision for Beta is missing.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied chart register to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Name the changed assumption and identify what must be rechecked before reusing the record. Use the edition register and printed legend to explain the inconsistency; request an applicable extract if necessary.

> **SCORING BOUNDARY**
>
> The task measures evidence versioning. Geographical familiarity should provide no scoring advantage. Require a causal explanation of why the original record may no longer apply. Accept more than one reasonable way to request a new check. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A04-01 · Equipment eligibility record

CURRENT-VERSUS-ARCHIVED

A training scenario inventory shows equipment marked available. A separate comparison page lists models and conditions that govern the simulated exercise.

CASE STIMULUS

Panel A is marked current in the interface. The register identifies A as an archived copy and B as the applicable revision. Both are accessible.

A: Equipment match / Orchard kit, rev 2, archived. B: rev 3. Register: rev 3 applies to this exercise.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied configuration sheet to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Explain which copy is applicable and cite the register entry and the revision label. Verify whether the configuration sheet covers the listed model and exercise before using availability as a proxy for eligibility.

> **SCORING BOUNDARY**
>
> Inventory availability and suitability are separate claims; require a statement of the missing relationship. Present both labels in equivalent positions. Mark a response incomplete if it selects B without identifying the evidence or if it confuses recency with applicability. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A04-02 · Equipment eligibility record

MISSING SCOPE

A training scenario inventory shows equipment marked available. A separate comparison page lists models and conditions that govern the simulated exercise.

CASE STIMULUS

The first source covers Object A only. The item asks about Object B, and the documentation for B is deliberately absent.

A: Equipment match / Orchard kit, equipment family=Alpha. Task: equipment family=Beta. Register: no Beta record supplied.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied configuration sheet to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Identify the scope mismatch, describe the missing evidence, and ask for a source explicitly covering B. Verify whether the configuration sheet covers the listed model and exercise before using availability as a proxy for eligibility.

> **SCORING BOUNDARY**
>
> Inventory availability and suitability are separate claims; require a statement of the missing relationship. A confident generalization to B is a consequential misconception in this exercise. Accept clear conditional explanations. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A04-03 · Equipment eligibility record

DUPLICATED SUMMARY

A training scenario inventory shows equipment marked available. A separate comparison page lists models and conditions that govern the simulated exercise.

CASE STIMULUS

Two interface summaries repeat the same claim. They resolve to a single upstream record released under an earlier scenario assumption.

Tile A and tile B cite Equipment match / Orchard kit, origin Q4. Q4 assumes training category=Alpha. Task fixes training category=Beta.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied configuration sheet to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Identify the common upstream source. Explain why repetition does not independently establish the claim under the current assumption. Verify whether the configuration sheet covers the listed model and exercise before using availability as a proxy for eligibility.

> **SCORING BOUNDARY**
>
> Inventory availability and suitability are separate claims; require a statement of the missing relationship. Require evidence that the two summaries share a dependency. Do not award full traceability for naming two tiles. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A04-04 · Equipment eligibility record

USABLE EVIDENCE

A training scenario inventory shows equipment marked available. A separate comparison page lists models and conditions that govern the simulated exercise.

CASE STIMULUS

The authoritative source explicitly covers the scenario, is within its validity window, and supports a single bounded conclusion.

A: Equipment match / Orchard kit, applies to this case. Register: A is valid for today's fictional session. A decision field: equipment eligibility confirmed.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied configuration sheet to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

State the limited conclusion with its source and explain why the evidence is sufficient for the question asked. Verify whether the configuration sheet covers the listed model and exercise before using availability as a proxy for eligibility.

> **SCORING BOUNDARY**
>
> Inventory availability and suitability are separate claims; require a statement of the missing relationship. This is an answerable case. Escalation alone is insufficient. Reward agreement limited to the question actually posed. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A04-05 · Equipment eligibility record

CONFLICTING RECORDS

A training scenario inventory shows equipment marked available. A separate comparison page lists models and conditions that govern the simulated exercise.

CASE STIMULUS

Two dated records make different assertions and the scenario pack provides no precedence rule. A separate note states which roles can resolve disagreements.

A: Equipment match / Orchard kit, equipment eligibility confirmed. B: equipment eligibility pending. No priority field. Role register: ask the independent scenario coordinator.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied configuration sheet to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Describe the conflict and identify a verifiable next action without selecting a winner on unsupported grounds. Verify whether the configuration sheet covers the listed model and exercise before using availability as a proxy for eligibility.

> **SCORING BOUNDARY**
>
> Inventory availability and suitability are separate claims; require a statement of the missing relationship. A response can give conditional implications for each record. It must avoid a final unsupported choice. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A04-06 · Equipment eligibility record

ASSUMPTION CHANGED

A training scenario inventory shows equipment marked available. A separate comparison page lists models and conditions that govern the simulated exercise.

CASE STIMULUS

The apparently useful record was generated under a named baseline assumption. The current scenario explicitly changes that assumption.

Record A: Equipment match / Orchard kit, training category=Alpha. Exercise overlay: training category=Beta. Revision for Beta is missing.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied configuration sheet to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Name the changed assumption and identify what must be rechecked before reusing the record. Verify whether the configuration sheet covers the listed model and exercise before using availability as a proxy for eligibility.

> **SCORING BOUNDARY**
>
> Inventory availability and suitability are separate claims; require a statement of the missing relationship. Require a causal explanation of why the original record may no longer apply. Accept more than one reasonable way to request a new check. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A05-01 · Mass worksheet metadata

CURRENT-VERSUS-ARCHIVED

Two fictional loading worksheets display identical output labels. One has a different unit convention in its legend and a separate assumptions note.

CASE STIMULUS

Panel A is marked current in the interface. The register identifies A as an archived copy and B as the applicable revision. Both are accessible.

A: Worksheet metadata / Jasper interface, rev 2, archived. B: rev 3. Register: rev 3 applies to this exercise.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied worksheet notes to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Explain which copy is applicable and cite the register entry and the revision label. Identify the unit convention and trace where it enters the displayed output; avoid carrying the result across an incompatible unit field.

> **SCORING BOUNDARY**
>
> All numeric values are placeholders. Score the recognition of a unit mismatch and the next check. Present both labels in equivalent positions. Mark a response incomplete if it selects B without identifying the evidence or if it confuses recency with applicability. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A05-02 · Mass worksheet metadata

MISSING SCOPE

Two fictional loading worksheets display identical output labels. One has a different unit convention in its legend and a separate assumptions note.

CASE STIMULUS

The first source covers Object A only. The item asks about Object B, and the documentation for B is deliberately absent.

A: Worksheet metadata / Jasper interface, unit type=Alpha. Task: unit type=Beta. Register: no Beta record supplied.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied worksheet notes to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Identify the scope mismatch, describe the missing evidence, and ask for a source explicitly covering B. Identify the unit convention and trace where it enters the displayed output; avoid carrying the result across an incompatible unit field.

> **SCORING BOUNDARY**
>
> All numeric values are placeholders. Score the recognition of a unit mismatch and the next check. A confident generalization to B is a consequential misconception in this exercise. Accept clear conditional explanations. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A05-03 · Mass worksheet metadata

DUPLICATED SUMMARY

Two fictional loading worksheets display identical output labels. One has a different unit convention in its legend and a separate assumptions note.

CASE STIMULUS

Two interface summaries repeat the same claim. They resolve to a single upstream record released under an earlier scenario assumption.

Tile A and tile B cite Worksheet metadata / Jasper interface, origin Q4. Q4 assumes label convention=Alpha. Task fixes label convention=Beta.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied worksheet notes to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Identify the common upstream source. Explain why repetition does not independently establish the claim under the current assumption. Identify the unit convention and trace where it enters the displayed output; avoid carrying the result across an incompatible unit field.

> **SCORING BOUNDARY**
>
> All numeric values are placeholders. Score the recognition of a unit mismatch and the next check. Require evidence that the two summaries share a dependency. Do not award full traceability for naming two tiles. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A05-04 · Mass worksheet metadata

USABLE EVIDENCE

Two fictional loading worksheets display identical output labels. One has a different unit convention in its legend and a separate assumptions note.

CASE STIMULUS

The authoritative source explicitly covers the scenario, is within its validity window, and supports a single bounded conclusion.

A: Worksheet metadata / Jasper interface, applies to this case. Register: A is valid for today's fictional session. A decision field: worksheet comparison confirmed.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied worksheet notes to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

State the limited conclusion with its source and explain why the evidence is sufficient for the question asked. Identify the unit convention and trace where it enters the displayed output; avoid carrying the result across an incompatible unit field.

> **SCORING BOUNDARY**
>
> All numeric values are placeholders. Score the recognition of a unit mismatch and the next check. This is an answerable case. Escalation alone is insufficient. Reward agreement limited to the question actually posed. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A05-05 · Mass worksheet metadata

CONFLICTING RECORDS

Two fictional loading worksheets display identical output labels. One has a different unit convention in its legend and a separate assumptions note.

CASE STIMULUS

Two dated records make different assertions and the scenario pack provides no precedence rule. A separate note states which roles can resolve disagreements.

A: Worksheet metadata / Jasper interface, worksheet comparison confirmed. B: worksheet comparison pending. No priority field. Role register: ask the independent scenario coordinator.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied worksheet notes to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Describe the conflict and identify a verifiable next action without selecting a winner on unsupported grounds. Identify the unit convention and trace where it enters the displayed output; avoid carrying the result across an incompatible unit field.

> **SCORING BOUNDARY**
>
> All numeric values are placeholders. Score the recognition of a unit mismatch and the next check. A response can give conditional implications for each record. It must avoid a final unsupported choice. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A05-06 · Mass worksheet metadata

ASSUMPTION CHANGED

Two fictional loading worksheets display identical output labels. One has a different unit convention in its legend and a separate assumptions note.

CASE STIMULUS

The apparently useful record was generated under a named baseline assumption. The current scenario explicitly changes that assumption.

Record A: Worksheet metadata / Jasper interface, label convention=Alpha. Exercise overlay: label convention=Beta. Revision for Beta is missing.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied worksheet notes to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Name the changed assumption and identify what must be rechecked before reusing the record. Identify the unit convention and trace where it enters the displayed output; avoid carrying the result across an incompatible unit field.

> **SCORING BOUNDARY**
>
> All numeric values are placeholders. Score the recognition of a unit mismatch and the next check. Require a causal explanation of why the original record may no longer apply. Accept more than one reasonable way to request a new check. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A06-01 · Planning source reconciliation

CURRENT-VERSUS-ARCHIVED

A route plan has a summary line marked current. The attached sources carry different effective windows for two parts of the example.

CASE STIMULUS

Panel A is marked current in the interface. The register identifies A as an archived copy and B as the applicable revision. Both are accessible.

A: Planning paperwork / Cobalt segment, rev 2, archived. B: rev 3. Register: rev 3 applies to this exercise.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied planning docket to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Explain which copy is applicable and cite the register entry and the revision label. Separate the time window that applies to each source and request confirmation when the source windows leave a gap.

> **SCORING BOUNDARY**
>
> Evidence coverage can vary across segments even when the summary looks internally consistent. Present both labels in equivalent positions. Mark a response incomplete if it selects B without identifying the evidence or if it confuses recency with applicability. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A06-02 · Planning source reconciliation

MISSING SCOPE

A route plan has a summary line marked current. The attached sources carry different effective windows for two parts of the example.

CASE STIMULUS

The first source covers Object A only. The item asks about Object B, and the documentation for B is deliberately absent.

A: Planning paperwork / Cobalt segment, segment identifier=Alpha. Task: segment identifier=Beta. Register: no Beta record supplied.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied planning docket to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Identify the scope mismatch, describe the missing evidence, and ask for a source explicitly covering B. Separate the time window that applies to each source and request confirmation when the source windows leave a gap.

> **SCORING BOUNDARY**
>
> Evidence coverage can vary across segments even when the summary looks internally consistent. A confident generalization to B is a consequential misconception in this exercise. Accept clear conditional explanations. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A06-03 · Planning source reconciliation

DUPLICATED SUMMARY

A route plan has a summary line marked current. The attached sources carry different effective windows for two parts of the example.

CASE STIMULUS

Two interface summaries repeat the same claim. They resolve to a single upstream record released under an earlier scenario assumption.

Tile A and tile B cite Planning paperwork / Cobalt segment, origin Q4. Q4 assumes time window=Alpha. Task fixes time window=Beta.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied planning docket to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Identify the common upstream source. Explain why repetition does not independently establish the claim under the current assumption. Separate the time window that applies to each source and request confirmation when the source windows leave a gap.

> **SCORING BOUNDARY**
>
> Evidence coverage can vary across segments even when the summary looks internally consistent. Require evidence that the two summaries share a dependency. Do not award full traceability for naming two tiles. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A06-04 · Planning source reconciliation

USABLE EVIDENCE

A route plan has a summary line marked current. The attached sources carry different effective windows for two parts of the example.

CASE STIMULUS

The authoritative source explicitly covers the scenario, is within its validity window, and supports a single bounded conclusion.

A: Planning paperwork / Cobalt segment, applies to this case. Register: A is valid for today's fictional session. A decision field: evidence coverage confirmed.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied planning docket to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

State the limited conclusion with its source and explain why the evidence is sufficient for the question asked. Separate the time window that applies to each source and request confirmation when the source windows leave a gap.

> **SCORING BOUNDARY**
>
> Evidence coverage can vary across segments even when the summary looks internally consistent. This is an answerable case. Escalation alone is insufficient. Reward agreement limited to the question actually posed. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A06-05 · Planning source reconciliation

CONFLICTING RECORDS

A route plan has a summary line marked current. The attached sources carry different effective windows for two parts of the example.

CASE STIMULUS

Two dated records make different assertions and the scenario pack provides no precedence rule. A separate note states which roles can resolve disagreements.

A: Planning paperwork / Cobalt segment, evidence coverage confirmed. B: evidence coverage pending. No priority field. Role register: ask the independent scenario coordinator.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied planning docket to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Describe the conflict and identify a verifiable next action without selecting a winner on unsupported grounds. Separate the time window that applies to each source and request confirmation when the source windows leave a gap.

> **SCORING BOUNDARY**
>
> Evidence coverage can vary across segments even when the summary looks internally consistent. A response can give conditional implications for each record. It must avoid a final unsupported choice. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A06-06 · Planning source reconciliation

ASSUMPTION CHANGED

A route plan has a summary line marked current. The attached sources carry different effective windows for two parts of the example.

CASE STIMULUS

The apparently useful record was generated under a named baseline assumption. The current scenario explicitly changes that assumption.

Record A: Planning paperwork / Cobalt segment, time window=Alpha. Exercise overlay: time window=Beta. Revision for Beta is missing.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied planning docket to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Name the changed assumption and identify what must be rechecked before reusing the record. Separate the time window that applies to each source and request confirmation when the source windows leave a gap.

> **SCORING BOUNDARY**
>
> Evidence coverage can vary across segments even when the summary looks internally consistent. Require a causal explanation of why the original record may no longer apply. Accept more than one reasonable way to request a new check. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A07-01 · Instructor handoff

CURRENT-VERSUS-ARCHIVED

A learner’s notes state they understand a topic. An instructor handoff card records the particular task and one unresolved uncertainty.

CASE STIMULUS

Panel A is marked current in the interface. The register identifies A as an archived copy and B as the applicable revision. Both are accessible.

A: Instructor handoff / Violet student record, rev 2, archived. B: rev 3. Register: rev 3 applies to this exercise.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied lesson handoff to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Explain which copy is applicable and cite the register entry and the revision label. Name the unresolved reasoning issue, show the supporting item, and propose a focused question for the instructor.

> **SCORING BOUNDARY**
>
> A concise, traceable question earns credit even if the learner uses different terminology from the model answer. Present both labels in equivalent positions. Mark a response incomplete if it selects B without identifying the evidence or if it confuses recency with applicability. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A07-02 · Instructor handoff

MISSING SCOPE

A learner’s notes state they understand a topic. An instructor handoff card records the particular task and one unresolved uncertainty.

CASE STIMULUS

The first source covers Object A only. The item asks about Object B, and the documentation for B is deliberately absent.

A: Instructor handoff / Violet student record, learning objective=Alpha. Task: learning objective=Beta. Register: no Beta record supplied.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied lesson handoff to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Identify the scope mismatch, describe the missing evidence, and ask for a source explicitly covering B. Name the unresolved reasoning issue, show the supporting item, and propose a focused question for the instructor.

> **SCORING BOUNDARY**
>
> A concise, traceable question earns credit even if the learner uses different terminology from the model answer. A confident generalization to B is a consequential misconception in this exercise. Accept clear conditional explanations. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A07-03 · Instructor handoff

DUPLICATED SUMMARY

A learner’s notes state they understand a topic. An instructor handoff card records the particular task and one unresolved uncertainty.

CASE STIMULUS

Two interface summaries repeat the same claim. They resolve to a single upstream record released under an earlier scenario assumption.

Tile A and tile B cite Instructor handoff / Violet student record, origin Q4. Q4 assumes study phase=Alpha. Task fixes study phase=Beta.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied lesson handoff to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Identify the common upstream source. Explain why repetition does not independently establish the claim under the current assumption. Name the unresolved reasoning issue, show the supporting item, and propose a focused question for the instructor.

> **SCORING BOUNDARY**
>
> A concise, traceable question earns credit even if the learner uses different terminology from the model answer. Require evidence that the two summaries share a dependency. Do not award full traceability for naming two tiles. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A07-04 · Instructor handoff

USABLE EVIDENCE

A learner’s notes state they understand a topic. An instructor handoff card records the particular task and one unresolved uncertainty.

CASE STIMULUS

The authoritative source explicitly covers the scenario, is within its validity window, and supports a single bounded conclusion.

A: Instructor handoff / Violet student record, applies to this case. Register: A is valid for today's fictional session. A decision field: next teaching action confirmed.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied lesson handoff to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

State the limited conclusion with its source and explain why the evidence is sufficient for the question asked. Name the unresolved reasoning issue, show the supporting item, and propose a focused question for the instructor.

> **SCORING BOUNDARY**
>
> A concise, traceable question earns credit even if the learner uses different terminology from the model answer. This is an answerable case. Escalation alone is insufficient. Reward agreement limited to the question actually posed. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A07-05 · Instructor handoff

CONFLICTING RECORDS

A learner’s notes state they understand a topic. An instructor handoff card records the particular task and one unresolved uncertainty.

CASE STIMULUS

Two dated records make different assertions and the scenario pack provides no precedence rule. A separate note states which roles can resolve disagreements.

A: Instructor handoff / Violet student record, next teaching action confirmed. B: next teaching action pending. No priority field. Role register: ask the independent scenario coordinator.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied lesson handoff to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Describe the conflict and identify a verifiable next action without selecting a winner on unsupported grounds. Name the unresolved reasoning issue, show the supporting item, and propose a focused question for the instructor.

> **SCORING BOUNDARY**
>
> A concise, traceable question earns credit even if the learner uses different terminology from the model answer. A response can give conditional implications for each record. It must avoid a final unsupported choice. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A07-06 · Instructor handoff

ASSUMPTION CHANGED

A learner’s notes state they understand a topic. An instructor handoff card records the particular task and one unresolved uncertainty.

CASE STIMULUS

The apparently useful record was generated under a named baseline assumption. The current scenario explicitly changes that assumption.

Record A: Instructor handoff / Violet student record, study phase=Alpha. Exercise overlay: study phase=Beta. Revision for Beta is missing.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied lesson handoff to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Name the changed assumption and identify what must be rechecked before reusing the record. Name the unresolved reasoning issue, show the supporting item, and propose a focused question for the instructor.

> **SCORING BOUNDARY**
>
> A concise, traceable question earns credit even if the learner uses different terminology from the model answer. Require a causal explanation of why the original record may no longer apply. Accept more than one reasonable way to request a new check. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A08-01 · Simulation assumption record

CURRENT-VERSUS-ARCHIVED

The training setup lists a default environment, while an instructor memo changes an assumption for one scenario only.

CASE STIMULUS

Panel A is marked current in the interface. The register identifies A as an archived copy and B as the applicable revision. Both are accessible.

A: Simulator setup / Meadow practice profile, rev 2, archived. B: rev 3. Register: rev 3 applies to this exercise.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied scenario assumptions to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Explain which copy is applicable and cite the register entry and the revision label. Check whether the change applies to the current scenario, then explain what judgment depends on it.

> **SCORING BOUNDARY**
>
> The task concerns simulator configuration provenance and requires no flight-control skill. Present both labels in equivalent positions. Mark a response incomplete if it selects B without identifying the evidence or if it confuses recency with applicability. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A08-02 · Simulation assumption record

MISSING SCOPE

The training setup lists a default environment, while an instructor memo changes an assumption for one scenario only.

CASE STIMULUS

The first source covers Object A only. The item asks about Object B, and the documentation for B is deliberately absent.

A: Simulator setup / Meadow practice profile, training profile=Alpha. Task: training profile=Beta. Register: no Beta record supplied.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied scenario assumptions to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Identify the scope mismatch, describe the missing evidence, and ask for a source explicitly covering B. Check whether the change applies to the current scenario, then explain what judgment depends on it.

> **SCORING BOUNDARY**
>
> The task concerns simulator configuration provenance and requires no flight-control skill. A confident generalization to B is a consequential misconception in this exercise. Accept clear conditional explanations. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A08-03 · Simulation assumption record

DUPLICATED SUMMARY

The training setup lists a default environment, while an instructor memo changes an assumption for one scenario only.

CASE STIMULUS

Two interface summaries repeat the same claim. They resolve to a single upstream record released under an earlier scenario assumption.

Tile A and tile B cite Simulator setup / Meadow practice profile, origin Q4. Q4 assumes environment flag=Alpha. Task fixes environment flag=Beta.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied scenario assumptions to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Identify the common upstream source. Explain why repetition does not independently establish the claim under the current assumption. Check whether the change applies to the current scenario, then explain what judgment depends on it.

> **SCORING BOUNDARY**
>
> The task concerns simulator configuration provenance and requires no flight-control skill. Require evidence that the two summaries share a dependency. Do not award full traceability for naming two tiles. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A08-04 · Simulation assumption record

USABLE EVIDENCE

The training setup lists a default environment, while an instructor memo changes an assumption for one scenario only.

CASE STIMULUS

The authoritative source explicitly covers the scenario, is within its validity window, and supports a single bounded conclusion.

A: Simulator setup / Meadow practice profile, applies to this case. Register: A is valid for today's fictional session. A decision field: baseline setup confirmed.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied scenario assumptions to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

State the limited conclusion with its source and explain why the evidence is sufficient for the question asked. Check whether the change applies to the current scenario, then explain what judgment depends on it.

> **SCORING BOUNDARY**
>
> The task concerns simulator configuration provenance and requires no flight-control skill. This is an answerable case. Escalation alone is insufficient. Reward agreement limited to the question actually posed. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A08-05 · Simulation assumption record

CONFLICTING RECORDS

The training setup lists a default environment, while an instructor memo changes an assumption for one scenario only.

CASE STIMULUS

Two dated records make different assertions and the scenario pack provides no precedence rule. A separate note states which roles can resolve disagreements.

A: Simulator setup / Meadow practice profile, baseline setup confirmed. B: baseline setup pending. No priority field. Role register: ask the independent scenario coordinator.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied scenario assumptions to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Describe the conflict and identify a verifiable next action without selecting a winner on unsupported grounds. Check whether the change applies to the current scenario, then explain what judgment depends on it.

> **SCORING BOUNDARY**
>
> The task concerns simulator configuration provenance and requires no flight-control skill. A response can give conditional implications for each record. It must avoid a final unsupported choice. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A08-06 · Simulation assumption record

ASSUMPTION CHANGED

The training setup lists a default environment, while an instructor memo changes an assumption for one scenario only.

CASE STIMULUS

The apparently useful record was generated under a named baseline assumption. The current scenario explicitly changes that assumption.

Record A: Simulator setup / Meadow practice profile, environment flag=Alpha. Exercise overlay: environment flag=Beta. Revision for Beta is missing.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied scenario assumptions to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Name the changed assumption and identify what must be rechecked before reusing the record. Check whether the change applies to the current scenario, then explain what judgment depends on it.

> **SCORING BOUNDARY**
>
> The task concerns simulator configuration provenance and requires no flight-control skill. Require a causal explanation of why the original record may no longer apply. Accept more than one reasonable way to request a new check. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A09-01 · Scheduling and permissions

CURRENT-VERSUS-ARCHIVED

A generic calendar shows an instructor or resource as available. A separate permissions record covers a narrower activity and time window.

CASE STIMULUS

Panel A is marked current in the interface. The register identifies A as an archived copy and B as the applicable revision. Both are accessible.

A: Scheduling availability / Amber resource slot, rev 2, archived. B: rev 3. Register: rev 3 applies to this exercise.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied permissions register to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Explain which copy is applicable and cite the register entry and the revision label. Separate scheduling from permission, and identify who can confirm the missing element in the fictional case.

> **SCORING BOUNDARY**
>
> Avoid scoring the choice of a particular staff role unless the provided documents define it. Present both labels in equivalent positions. Mark a response incomplete if it selects B without identifying the evidence or if it confuses recency with applicability. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A09-02 · Scheduling and permissions

MISSING SCOPE

A generic calendar shows an instructor or resource as available. A separate permissions record covers a narrower activity and time window.

CASE STIMULUS

The first source covers Object A only. The item asks about Object B, and the documentation for B is deliberately absent.

A: Scheduling availability / Amber resource slot, booking class=Alpha. Task: booking class=Beta. Register: no Beta record supplied.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied permissions register to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Identify the scope mismatch, describe the missing evidence, and ask for a source explicitly covering B. Separate scheduling from permission, and identify who can confirm the missing element in the fictional case.

> **SCORING BOUNDARY**
>
> Avoid scoring the choice of a particular staff role unless the provided documents define it. A confident generalization to B is a consequential misconception in this exercise. Accept clear conditional explanations. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A09-03 · Scheduling and permissions

DUPLICATED SUMMARY

A generic calendar shows an instructor or resource as available. A separate permissions record covers a narrower activity and time window.

CASE STIMULUS

Two interface summaries repeat the same claim. They resolve to a single upstream record released under an earlier scenario assumption.

Tile A and tile B cite Scheduling availability / Amber resource slot, origin Q4. Q4 assumes schedule revision=Alpha. Task fixes schedule revision=Beta.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied permissions register to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Identify the common upstream source. Explain why repetition does not independently establish the claim under the current assumption. Separate scheduling from permission, and identify who can confirm the missing element in the fictional case.

> **SCORING BOUNDARY**
>
> Avoid scoring the choice of a particular staff role unless the provided documents define it. Require evidence that the two summaries share a dependency. Do not award full traceability for naming two tiles. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A09-04 · Scheduling and permissions

USABLE EVIDENCE

A generic calendar shows an instructor or resource as available. A separate permissions record covers a narrower activity and time window.

CASE STIMULUS

The authoritative source explicitly covers the scenario, is within its validity window, and supports a single bounded conclusion.

A: Scheduling availability / Amber resource slot, applies to this case. Register: A is valid for today's fictional session. A decision field: authorization status confirmed.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied permissions register to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

State the limited conclusion with its source and explain why the evidence is sufficient for the question asked. Separate scheduling from permission, and identify who can confirm the missing element in the fictional case.

> **SCORING BOUNDARY**
>
> Avoid scoring the choice of a particular staff role unless the provided documents define it. This is an answerable case. Escalation alone is insufficient. Reward agreement limited to the question actually posed. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A09-05 · Scheduling and permissions

CONFLICTING RECORDS

A generic calendar shows an instructor or resource as available. A separate permissions record covers a narrower activity and time window.

CASE STIMULUS

Two dated records make different assertions and the scenario pack provides no precedence rule. A separate note states which roles can resolve disagreements.

A: Scheduling availability / Amber resource slot, authorization status confirmed. B: authorization status pending. No priority field. Role register: ask the independent scenario coordinator.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied permissions register to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Describe the conflict and identify a verifiable next action without selecting a winner on unsupported grounds. Separate scheduling from permission, and identify who can confirm the missing element in the fictional case.

> **SCORING BOUNDARY**
>
> Avoid scoring the choice of a particular staff role unless the provided documents define it. A response can give conditional implications for each record. It must avoid a final unsupported choice. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A09-06 · Scheduling and permissions

ASSUMPTION CHANGED

A generic calendar shows an instructor or resource as available. A separate permissions record covers a narrower activity and time window.

CASE STIMULUS

The apparently useful record was generated under a named baseline assumption. The current scenario explicitly changes that assumption.

Record A: Scheduling availability / Amber resource slot, schedule revision=Alpha. Exercise overlay: schedule revision=Beta. Revision for Beta is missing.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied permissions register to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Name the changed assumption and identify what must be rechecked before reusing the record. Separate scheduling from permission, and identify who can confirm the missing element in the fictional case.

> **SCORING BOUNDARY**
>
> Avoid scoring the choice of a particular staff role unless the provided documents define it. Require a causal explanation of why the original record may no longer apply. Accept more than one reasonable way to request a new check. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A10-01 · Logbook correction chain

CURRENT-VERSUS-ARCHIVED

An activity is visible in a fictional logbook. A correction notice supersedes one field while the remaining fields are retained in the archived entry.

CASE STIMULUS

Panel A is marked current in the interface. The register identifies A as an archived copy and B as the applicable revision. Both are accessible.

A: Logbook updates / Birch correction, rev 2, archived. B: rev 3. Register: rev 3 applies to this exercise.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied correction notice to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Explain which copy is applicable and cite the register entry and the revision label. Trace exactly which field changed, identify the current record, and preserve the reason for the correction.

> **SCORING BOUNDARY**
>
> The construct is change tracking, not sign-off authority or recordkeeping advice. Present both labels in equivalent positions. Mark a response incomplete if it selects B without identifying the evidence or if it confuses recency with applicability. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A10-02 · Logbook correction chain

MISSING SCOPE

An activity is visible in a fictional logbook. A correction notice supersedes one field while the remaining fields are retained in the archived entry.

CASE STIMULUS

The first source covers Object A only. The item asks about Object B, and the documentation for B is deliberately absent.

A: Logbook updates / Birch correction, entry identifier=Alpha. Task: entry identifier=Beta. Register: no Beta record supplied.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied correction notice to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Identify the scope mismatch, describe the missing evidence, and ask for a source explicitly covering B. Trace exactly which field changed, identify the current record, and preserve the reason for the correction.

> **SCORING BOUNDARY**
>
> The construct is change tracking, not sign-off authority or recordkeeping advice. A confident generalization to B is a consequential misconception in this exercise. Accept clear conditional explanations. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A10-03 · Logbook correction chain

DUPLICATED SUMMARY

An activity is visible in a fictional logbook. A correction notice supersedes one field while the remaining fields are retained in the archived entry.

CASE STIMULUS

Two interface summaries repeat the same claim. They resolve to a single upstream record released under an earlier scenario assumption.

Tile A and tile B cite Logbook updates / Birch correction, origin Q4. Q4 assumes changed field=Alpha. Task fixes changed field=Beta.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied correction notice to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Identify the common upstream source. Explain why repetition does not independently establish the claim under the current assumption. Trace exactly which field changed, identify the current record, and preserve the reason for the correction.

> **SCORING BOUNDARY**
>
> The construct is change tracking, not sign-off authority or recordkeeping advice. Require evidence that the two summaries share a dependency. Do not award full traceability for naming two tiles. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A10-04 · Logbook correction chain

USABLE EVIDENCE

An activity is visible in a fictional logbook. A correction notice supersedes one field while the remaining fields are retained in the archived entry.

CASE STIMULUS

The authoritative source explicitly covers the scenario, is within its validity window, and supports a single bounded conclusion.

A: Logbook updates / Birch correction, applies to this case. Register: A is valid for today's fictional session. A decision field: amended field confirmed.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied correction notice to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

State the limited conclusion with its source and explain why the evidence is sufficient for the question asked. Trace exactly which field changed, identify the current record, and preserve the reason for the correction.

> **SCORING BOUNDARY**
>
> The construct is change tracking, not sign-off authority or recordkeeping advice. This is an answerable case. Escalation alone is insufficient. Reward agreement limited to the question actually posed. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A10-05 · Logbook correction chain

CONFLICTING RECORDS

An activity is visible in a fictional logbook. A correction notice supersedes one field while the remaining fields are retained in the archived entry.

CASE STIMULUS

Two dated records make different assertions and the scenario pack provides no precedence rule. A separate note states which roles can resolve disagreements.

A: Logbook updates / Birch correction, amended field confirmed. B: amended field pending. No priority field. Role register: ask the independent scenario coordinator.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied correction notice to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Describe the conflict and identify a verifiable next action without selecting a winner on unsupported grounds. Trace exactly which field changed, identify the current record, and preserve the reason for the correction.

> **SCORING BOUNDARY**
>
> The construct is change tracking, not sign-off authority or recordkeeping advice. A response can give conditional implications for each record. It must avoid a final unsupported choice. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A10-06 · Logbook correction chain

ASSUMPTION CHANGED

An activity is visible in a fictional logbook. A correction notice supersedes one field while the remaining fields are retained in the archived entry.

CASE STIMULUS

The apparently useful record was generated under a named baseline assumption. The current scenario explicitly changes that assumption.

Record A: Logbook updates / Birch correction, changed field=Alpha. Exercise overlay: changed field=Beta. Revision for Beta is missing.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied correction notice to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Name the changed assumption and identify what must be rechecked before reusing the record. Trace exactly which field changed, identify the current record, and preserve the reason for the correction.

> **SCORING BOUNDARY**
>
> The construct is change tracking, not sign-off authority or recordkeeping advice. Require a causal explanation of why the original record may no longer apply. Accept more than one reasonable way to request a new check. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A11-01 · Shared-responsibility disclosure

CURRENT-VERSUS-ARCHIVED

A simulated marketplace lists an activity with a confidence-inspiring summary. Its role card allocates separate decisions to the participants.

CASE STIMULUS

Panel A is marked current in the interface. The register identifies A as an archived copy and B as the applicable revision. Both are accessible.

A: Role disclosure / Cedar shared exercise, rev 2, archived. B: rev 3. Register: rev 3 applies to this exercise.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied role card to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Explain which copy is applicable and cite the register entry and the revision label. Identify the decision that still belongs to a participant and link that responsibility to the specific supplied record.

> **SCORING BOUNDARY**
>
> Do not treat a product claim as proof of regulatory compliance. This is a scenario about comprehension. Present both labels in equivalent positions. Mark a response incomplete if it selects B without identifying the evidence or if it confuses recency with applicability. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A11-02 · Shared-responsibility disclosure

MISSING SCOPE

A simulated marketplace lists an activity with a confidence-inspiring summary. Its role card allocates separate decisions to the participants.

CASE STIMULUS

The first source covers Object A only. The item asks about Object B, and the documentation for B is deliberately absent.

A: Role disclosure / Cedar shared exercise, role identifier=Alpha. Task: role identifier=Beta. Register: no Beta record supplied.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied role card to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Identify the scope mismatch, describe the missing evidence, and ask for a source explicitly covering B. Identify the decision that still belongs to a participant and link that responsibility to the specific supplied record.

> **SCORING BOUNDARY**
>
> Do not treat a product claim as proof of regulatory compliance. This is a scenario about comprehension. A confident generalization to B is a consequential misconception in this exercise. Accept clear conditional explanations. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A11-03 · Shared-responsibility disclosure

DUPLICATED SUMMARY

A simulated marketplace lists an activity with a confidence-inspiring summary. Its role card allocates separate decisions to the participants.

CASE STIMULUS

Two interface summaries repeat the same claim. They resolve to a single upstream record released under an earlier scenario assumption.

Tile A and tile B cite Role disclosure / Cedar shared exercise, origin Q4. Q4 assumes participant type=Alpha. Task fixes participant type=Beta.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied role card to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Identify the common upstream source. Explain why repetition does not independently establish the claim under the current assumption. Identify the decision that still belongs to a participant and link that responsibility to the specific supplied record.

> **SCORING BOUNDARY**
>
> Do not treat a product claim as proof of regulatory compliance. This is a scenario about comprehension. Require evidence that the two summaries share a dependency. Do not award full traceability for naming two tiles. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A11-04 · Shared-responsibility disclosure

USABLE EVIDENCE

A simulated marketplace lists an activity with a confidence-inspiring summary. Its role card allocates separate decisions to the participants.

CASE STIMULUS

The authoritative source explicitly covers the scenario, is within its validity window, and supports a single bounded conclusion.

A: Role disclosure / Cedar shared exercise, applies to this case. Register: A is valid for today's fictional session. A decision field: assigned role confirmed.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied role card to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

State the limited conclusion with its source and explain why the evidence is sufficient for the question asked. Identify the decision that still belongs to a participant and link that responsibility to the specific supplied record.

> **SCORING BOUNDARY**
>
> Do not treat a product claim as proof of regulatory compliance. This is a scenario about comprehension. This is an answerable case. Escalation alone is insufficient. Reward agreement limited to the question actually posed. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A11-05 · Shared-responsibility disclosure

CONFLICTING RECORDS

A simulated marketplace lists an activity with a confidence-inspiring summary. Its role card allocates separate decisions to the participants.

CASE STIMULUS

Two dated records make different assertions and the scenario pack provides no precedence rule. A separate note states which roles can resolve disagreements.

A: Role disclosure / Cedar shared exercise, assigned role confirmed. B: assigned role pending. No priority field. Role register: ask the independent scenario coordinator.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied role card to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Describe the conflict and identify a verifiable next action without selecting a winner on unsupported grounds. Identify the decision that still belongs to a participant and link that responsibility to the specific supplied record.

> **SCORING BOUNDARY**
>
> Do not treat a product claim as proof of regulatory compliance. This is a scenario about comprehension. A response can give conditional implications for each record. It must avoid a final unsupported choice. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A11-06 · Shared-responsibility disclosure

ASSUMPTION CHANGED

A simulated marketplace lists an activity with a confidence-inspiring summary. Its role card allocates separate decisions to the participants.

CASE STIMULUS

The apparently useful record was generated under a named baseline assumption. The current scenario explicitly changes that assumption.

Record A: Role disclosure / Cedar shared exercise, participant type=Alpha. Exercise overlay: participant type=Beta. Revision for Beta is missing.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied role card to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Name the changed assumption and identify what must be rechecked before reusing the record. Identify the decision that still belongs to a participant and link that responsibility to the specific supplied record.

> **SCORING BOUNDARY**
>
> Do not treat a product claim as proof of regulatory compliance. This is a scenario about comprehension. Require a causal explanation of why the original record may no longer apply. Accept more than one reasonable way to request a new check. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A12-01 · Source-grounded feedback audit

CURRENT-VERSUS-ARCHIVED

An instructional response has a confident conclusion and a reference label. The supplied source excerpt supports a narrower statement than the conclusion.

CASE STIMULUS

Panel A is marked current in the interface. The register identifies A as an archived copy and B as the applicable revision. Both are accessible.

A: Feedback claims / Quartz source packet, rev 2, archived. B: rev 3. Register: rev 3 applies to this exercise.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied source excerpt to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Explain which copy is applicable and cite the register entry and the revision label. Compare the cited material to the claim; state which part is supported and what would require another source.

> **SCORING BOUNDARY**
>
> Credit useful partial support and explicit limits. A fluent explanation can still make an unsupported inference. Present both labels in equivalent positions. Mark a response incomplete if it selects B without identifying the evidence or if it confuses recency with applicability. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A12-02 · Source-grounded feedback audit

MISSING SCOPE

An instructional response has a confident conclusion and a reference label. The supplied source excerpt supports a narrower statement than the conclusion.

CASE STIMULUS

The first source covers Object A only. The item asks about Object B, and the documentation for B is deliberately absent.

A: Feedback claims / Quartz source packet, authoritative section=Alpha. Task: authoritative section=Beta. Register: no Beta record supplied.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied source excerpt to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Identify the scope mismatch, describe the missing evidence, and ask for a source explicitly covering B. Compare the cited material to the claim; state which part is supported and what would require another source.

> **SCORING BOUNDARY**
>
> Credit useful partial support and explicit limits. A fluent explanation can still make an unsupported inference. A confident generalization to B is a consequential misconception in this exercise. Accept clear conditional explanations. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A12-03 · Source-grounded feedback audit

DUPLICATED SUMMARY

An instructional response has a confident conclusion and a reference label. The supplied source excerpt supports a narrower statement than the conclusion.

CASE STIMULUS

Two interface summaries repeat the same claim. They resolve to a single upstream record released under an earlier scenario assumption.

Tile A and tile B cite Feedback claims / Quartz source packet, origin Q4. Q4 assumes evidence type=Alpha. Task fixes evidence type=Beta.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied source excerpt to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Identify the common upstream source. Explain why repetition does not independently establish the claim under the current assumption. Compare the cited material to the claim; state which part is supported and what would require another source.

> **SCORING BOUNDARY**
>
> Credit useful partial support and explicit limits. A fluent explanation can still make an unsupported inference. Require evidence that the two summaries share a dependency. Do not award full traceability for naming two tiles. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A12-04 · Source-grounded feedback audit

USABLE EVIDENCE

An instructional response has a confident conclusion and a reference label. The supplied source excerpt supports a narrower statement than the conclusion.

CASE STIMULUS

The authoritative source explicitly covers the scenario, is within its validity window, and supports a single bounded conclusion.

A: Feedback claims / Quartz source packet, applies to this case. Register: A is valid for today's fictional session. A decision field: claim boundary confirmed.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied source excerpt to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

State the limited conclusion with its source and explain why the evidence is sufficient for the question asked. Compare the cited material to the claim; state which part is supported and what would require another source.

> **SCORING BOUNDARY**
>
> Credit useful partial support and explicit limits. A fluent explanation can still make an unsupported inference. This is an answerable case. Escalation alone is insufficient. Reward agreement limited to the question actually posed. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A12-05 · Source-grounded feedback audit

CONFLICTING RECORDS

An instructional response has a confident conclusion and a reference label. The supplied source excerpt supports a narrower statement than the conclusion.

CASE STIMULUS

Two dated records make different assertions and the scenario pack provides no precedence rule. A separate note states which roles can resolve disagreements.

A: Feedback claims / Quartz source packet, claim boundary confirmed. B: claim boundary pending. No priority field. Role register: ask the independent scenario coordinator.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied source excerpt to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Describe the conflict and identify a verifiable next action without selecting a winner on unsupported grounds. Compare the cited material to the claim; state which part is supported and what would require another source.

> **SCORING BOUNDARY**
>
> Credit useful partial support and explicit limits. A fluent explanation can still make an unsupported inference. A response can give conditional implications for each record. It must avoid a final unsupported choice. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

A12-06 · Source-grounded feedback audit

ASSUMPTION CHANGED

An instructional response has a confident conclusion and a reference label. The supplied source excerpt supports a narrower statement than the conclusion.

CASE STIMULUS

The apparently useful record was generated under a named baseline assumption. The current scenario explicitly changes that assumption.

Record A: Feedback claims / Quartz source packet, evidence type=Alpha. Exercise overlay: evidence type=Beta. Revision for Beta is missing.

LEARNER TASK

Give a short explanation identifying the issue and its implication. Use the supplied source excerpt to justify which source you checked and which action would verify the record.

INSTRUCTOR REFERENCE

Name the changed assumption and identify what must be rechecked before reusing the record. Compare the cited material to the claim; state which part is supported and what would require another source.

> **SCORING BOUNDARY**
>
> Credit useful partial support and explicit limits. A fluent explanation can still make an unsupported inference. Require a causal explanation of why the original record may no longer apply. Accept more than one reasonable way to request a new check. Apply the four 0-2 rubric dimensions. Use an independent protected bank for confirmatory assessment.

B01 · Site authorization worksheet

RECORD OWNER AND HANDOFF

Site authorizer to coordinator

Confirm that the location, people, recruitment process and software access match the approved protocol before scheduling.

Study setting and independent point of contact

scope of permitted activity

governance approvals and expiration

access revocation owner

INDEPENDENT REVIEW

Record reviewer role, date, document version, decision, and any issue that needs follow-up. Attach approved evidence separately when it contains restricted information.

B02 · Participant information outline

RECORD OWNER AND HANDOFF

Coordinator to independent ethics reviewer

Use the approved participant language in the study. This worksheet records review items rather than acting as a consent form.

Purpose and time commitments

assignment to two feedback formats

voluntary participation

data use

withdrawal and complaints

INDEPENDENT REVIEW

Record reviewer role, date, document version, decision, and any issue that needs follow-up. Attach approved evidence separately when it contains restricted information.

B03 · Consent version register

RECORD OWNER AND HANDOFF

Coordinator to data steward

Store signed or attributable material outside the analytic dataset. The trial ledger carries only a consent status.

Consent version and language	
approval date	
mode of collection	
accessible alternative	
discrepancy route	

INDEPENDENT REVIEW

Record reviewer role, date, document version, decision, and any issue that needs follow-up. Attach approved evidence separately when it contains restricted information.

B04 · Recruitment screening record

RECORD OWNER AND HANDOFF

Recruiter to allocator

Apply the same eligibility questions before condition assignment. Keep recruitment information from the founder’s commercial team.

Eligibility conditions
prior exposure
provider and stratum
reason for screen-out
route to ask questions

INDEPENDENT REVIEW

Record reviewer role, date, document version, decision, and any issue that needs follow-up. Attach approved evidence separately when it contains restricted information.

B05 · Allocation concealment record

RECORD OWNER AND HANDOFF

Independent allocator to coordinator

Log deviations at discovery, including accidental early disclosure, with the affected scope and mitigation.

Sequence custodian

stratum

time of assignment

exception identifier

evidence of masking

INDEPENDENT REVIEW

Record reviewer role, date, document version, decision, and any issue that needs follow-up. Attach approved evidence separately when it contains restricted information.

B06 · Intervention description record

RECORD OWNER AND HANDOFF

Instructor to monitor

Record enough detail to distinguish the assigned comparison and the intervention participants received.

Objective

versioned materials

instructor role

procedure and duration

permitted adaptation

actual delivery

INDEPENDENT REVIEW

Record reviewer role, date, document version, decision, and any issue that needs follow-up. Attach approved evidence separately when it contains restricted information.

B07 · Tutor configuration card

RECORD OWNER AND HANDOFF

Engineer to benchmark custodian

Archive time-stamped configurations and comparator outputs. Record cases where provider versioning is incomplete.

Model and date

acquisition method

prompt and retrieval versions

reference hashes

refusal and referral behavior

INDEPENDENT REVIEW

Record reviewer role, date, document version, decision, and any issue that needs follow-up. Attach approved evidence separately when it contains restricted information.

B08 · Standard feedback comparator card

RECORD OWNER AND HANDOFF

Independent instructor to evaluator

Match the intervention on time, source access, and lesson objective. Review the static text against the same standard used for the tutor.

Course objective

author/approval role

source pack

feedback rules

change history

INDEPENDENT REVIEW

Record reviewer role, date, document version, decision, and any issue that needs follow-up. Attach approved evidence separately when it contains restricted information.

B09 · Readiness gate decision

RECORD OWNER AND HANDOFF

Independent panel to committee

Fix the gate before scoring. Repair and use a fresh protected holdout if the gate is missed.

Consequential error definition

threshold and interval method

adjudication completed

decision and rationale

INDEPENDENT REVIEW

Record reviewer role, date, document version, decision, and any issue that needs follow-up. Attach approved evidence separately when it contains restricted information.

B10 · Lesson exposure form

RECORD OWNER AND HANDOFF

Session monitor to coordinator

Record implementation details without removing participants from assigned-arm comparisons.

Start/end window

delivery status

activity completion

permitted instructor involvement

technology interruption

INDEPENDENT REVIEW

Record reviewer role, date, document version, decision, and any issue that needs follow-up. Attach approved evidence separately when it contains restricted information.

B11 · Baseline assessment card

RECORD OWNER AND HANDOFF

Assessor to allocator

Complete baseline scoring before assignment is disclosed. Document accommodation and preserve identical comparison rules.

Form ID and version

previous exposure

masked identifier

approved aid

completeness

INDEPENDENT REVIEW

Record reviewer role, date, document version, decision, and any issue that needs follow-up. Attach approved evidence separately when it contains restricted information.

B12 · Delayed assessment session log

RECORD OWNER AND HANDOFF

Assessor to data steward

Track timing relative to instruction. A late session should remain distinguishable from an absent or refused session.

Target date and allowed window

form and device

allowed sources

interruption and reschedule status

INDEPENDENT REVIEW

Record reviewer role, date, document version, decision, and any issue that needs follow-up. Attach approved evidence separately when it contains restricted information.

B13 · Transfer simulation log

RECORD OWNER AND HANDOFF

Independent assessor to replication lead

Use fictional inputs and identical support across arms. Record software faults separately from participant behavior.

Sandbox version

task order

static references

tutor unavailable

connection and reset integrity

INDEPENDENT REVIEW

Record reviewer role, date, document version, decision, and any issue that needs follow-up. Attach approved evidence separately when it contains restricted information.

B14 · Masked scoring sheet

RECORD OWNER AND HANDOFF

Scorer to adjudicator

Use the scored answer and supplied evidence, including a defensible alternative. Preserve the first score for agreement analysis.

Recognition score

assumption score

source score

action score

evidence note and masking disclosure flag

INDEPENDENT REVIEW

Record reviewer role, date, document version, decision, and any issue that needs follow-up. Attach approved evidence separately when it contains restricted information.

B15 · Adjudication report

RECORD OWNER AND HANDOFF

Third reviewer to steward

Resolve disagreements independently of study arm. Record ambiguity before deciding whether the item can contribute to analysis.

Rater scores

source inspected

disputed criterion

resolution and reviewer

item-level concern

INDEPENDENT REVIEW

Record reviewer role, date, document version, decision, and any issue that needs follow-up. Attach approved evidence separately when it contains restricted information.

B16 · Workload and access debrief

RECORD OWNER AND HANDOFF

Participant-facing coordinator to analyst

Treat implementation as a secondary outcome. Avoid conditioning primary effectiveness estimates on convenience or satisfaction.

Perceived effort

clarity

barriers

unplanned help

discomfort or early exit

INDEPENDENT REVIEW

Record reviewer role, date, document version, decision, and any issue that needs follow-up. Attach approved evidence separately when it contains restricted information.

B17 · Protocol deviation form

RECORD OWNER AND HANDOFF

Monitor to principal investigator

Write the account before examining unmasked estimates when practicable. Preserve previous versions and the rationale.

What changed

discoverer

when and how detected

affected arms or forms

follow-up and disclosure

INDEPENDENT REVIEW

Record reviewer role, date, document version, decision, and any issue that needs follow-up. Attach approved evidence separately when it contains restricted information.

B18 · Adverse-event and pause card

RECORD OWNER AND HANDOFF

Coordinator to oversight body

Follow the institution-approved reporting pathway and note whether procedures differ across conditions.

Nature of event

immediate response

affected process

independent decision

route to restart or stop

INDEPENDENT REVIEW

Record reviewer role, date, document version, decision, and any issue that needs follow-up. Attach approved evidence separately when it contains restricted information.

B19 · Missing follow-up worksheet

RECORD OWNER AND HANDOFF

Coordinator to analyst

Use the same contact strategy in each arm. Supply categories needed for the sensitivity analysis while protecting identifying details.

Due date

attempted contacts

reason category

late window or withdrawal

permission for follow-up

INDEPENDENT REVIEW

Record reviewer role, date, document version, decision, and any issue that needs follow-up. Attach approved evidence separately when it contains restricted information.

B20 · Independent analysis handoff

RECORD OWNER AND HANDOFF

Masked analyst to independent checker

Rerun critical estimates from a clean environment and reconcile discrepancies before the arm mapping is opened.

Frozen inputs and hashes
dated script
coded arm mapping custodian
log of reruns
output approval

INDEPENDENT REVIEW

Record reviewer role, date, document version, decision, and any issue that needs follow-up. Attach approved evidence separately when it contains restricted information.

B21 · AI assistance disclosure log

RECORD OWNER AND HANDOFF

Candidate to supervisor

Log research assistance separately from the learner’s exposure to the experimental tutor.

Tool version and date

reason for use

affected artifact

data category

independent verification and author sign-off

INDEPENDENT REVIEW

Record reviewer role, date, document version, decision, and any issue that needs follow-up. Attach approved evidence separately when it contains restricted information.

B22 · Release and licensing checklist

RECORD OWNER AND HANDOFF

Steward to release approver

Separate original examples from excerpts, dependencies, and restricted data. Release a synthetic substitute when needed.

Owner and license

consent coverage

redaction

prohibited assets

approved publication destination

INDEPENDENT REVIEW

Record reviewer role, date, document version, decision, and any issue that needs follow-up. Attach approved evidence separately when it contains restricted information.

B23 · Committee decision cover sheet

RECORD OWNER AND HANDOFF

Committee to candidate

Record the decision and the exact document it applies to. A proposal may return for revision before recruitment.

Version reviewed

methods clearance

outstanding ethics or data conditions

expected evidence

decision date

INDEPENDENT REVIEW

Record reviewer role, date, document version, decision, and any issue that needs follow-up. Attach approved evidence separately when it contains restricted information.

B24 · Defense and deposit checklist

RECORD OWNER AND HANDOFF

Candidate to committee and Registrar

A signed approval and designated deposit channel determine archival status. Record where the approved materials can be found.

Primary evidence summary

source and code manifests

disclosures

permitted attachments

deposit receipt

INDEPENDENT REVIEW

Record reviewer role, date, document version, decision, and any issue that needs follow-up. Attach approved evidence separately when it contains restricted information.

C01 · Participant and enrollment records

DATA FIELDS

participant_id	pseudonymous stable text
site_code	categorical
eligibility	boolean
consent_version	text
baseline_complete	boolean

VALIDATION RULE

Randomization can occur only after eligible and consented status is documented. Separate linking keys from research files.

USE IN INTERPRETATION

When exporting a synthetic dataset use obviously non-personal identifiers. Preserve enrollment, consent, randomization and completion as distinct events.

> **REVIEW RECORD**
>
> Record the schema version and its owner. Explain any deviations from these proposed fields in the released codebook and replication notes.

C02 · Assignment and masking

DATA FIELDS

participant_id	foreign key
stratum	text
arm_code	categorical
allocation_timestamp	sortable timestamp
sequence_version	text

VALIDATION RULE

Arm labels in the analytic file are opaque until the primary output is locked. The sequence is held by an independent custodian.

USE IN INTERPRETATION

Log accidental unmasking with time, exposure and resolution. Do not silently replace assignment with actual exposure.

> **REVIEW RECORD**
>
> Record the schema version and its owner. Explain any deviations from these proposed fields in the released codebook and replication notes.

C03 · Intervention and comparator versions

DATA FIELDS

intervention_id	stable text
course_revision	text
tool_version	text
prompt_revision	text
source_digest	hash
comparator_revision	text

VALIDATION RULE

Every session links to exactly one intervention configuration. File unknown vendor model details as unknown.

USE IN INTERPRETATION

A binding between prompt, sources, and code is more informative than a marketing name. Distinguish a fix from a scheduled update.

> **REVIEW RECORD**
>
> Record the schema version and its owner. Explain any deviations from these proposed fields in the released codebook and replication notes.

C04 · Scenario and form manifest

DATA FIELDS

case_id	unique text
family	categorical
partition	categorical
form_version	text
stimulus_digest	hash
review_status	categorical

VALIDATION RULE

Holdouts are separate from training materials. Published examples are never confirmatory assessment items.

USE IN INTERPRETATION

Review overlap at the level of underlying evidence relationships. Test whether rotated forms measure the same construct.

> **REVIEW RECORD**
>
> Record the schema version and its owner. Explain any deviations from these proposed fields in the released codebook and replication notes.

C05 · Session and device exposure

DATA FIELDS

session_id	unique text
planned_window	timestamp range
device_class	categorical
exposure_minutes	nonnegative
interruption_code	categorical

VALIDATION RULE

Software interruptions are not learner errors. Label any alternate mode or later reschedule.

USE IN INTERPRETATION

Monitor intervention fidelity and report adverse events with denominators. Avoid using exposure minutes as a primary exclusion rule.

> **REVIEW RECORD**
>
> Record the schema version and its owner. Explain any deviations from these proposed fields in the released codebook and replication notes.

C06 · Item-level assessment

DATA FIELDS

participant_id	foreign key
item_id	foreign key
visit	categorical
response_id	text
independent_score_version	text

VALIDATION RULE

The raw written answer and the scoring record have a controlled relationship. Restrict identity access.

USE IN INTERPRETATION

Any permission to share response extracts is evaluated separately from permission to report aggregate data.

> **REVIEW RECORD**
>
> Record the schema version and its owner. Explain any deviations from these proposed fields in the released codebook and replication notes.

C07 · Rubric and adjudication

DATA FIELDS

recognition, assumption, traceability, action	integers 0-2
rater_id	pseudonymous text
rationale_id	text
rating_stage	categorical

VALIDATION RULE

Preserve original independent scores before adjudication. An adjudicated score has a recorded rationale.

USE IN INTERPRETATION

Distinct missing codes prevent absence from being confused with zero. Use a stable item key to reproduce disagreement checks.

> **REVIEW RECORD**
>
> Record the schema version and its owner. Explain any deviations from these proposed fields in the released codebook and replication notes.

C08 · Outcomes and visit windows

DATA FIELDS

visit_id	text
relative_day	integer
score_pct	numeric
contributing_items	integer
optional_workload	numeric

VALIDATION RULE

Keep day-14 assessment and transfer scores in distinct records. Analyze only according to prespecified outcome rules.

USE IN INTERPRETATION

Report visit timing, score distribution, performance interval, and planned sensitivity checks.

> **REVIEW RECORD**
>
> Record the schema version and its owner. Explain any deviations from these proposed fields in the released codebook and replication notes.

C09 · Missingness and deviations

DATA FIELDS

reason	categorical
downtime	boolean
planned_measure	text
exception_id	text
severity	categorical

VALIDATION RULE

Missing data remain missing. Operational deviations can affect completeness without being exclusion criteria.

USE IN INTERPRETATION

Prespecify robustness checks and state how differential missingness affects interpretation.

> **REVIEW RECORD**
>
> Record the schema version and its owner. Explain any deviations from these proposed fields in the released codebook and replication notes.

C10 · Analysis manifest and checks

DATA FIELDS

dataset_hash	text
script_hash	text
generated_at	timestamp
lock_version	text
reviewer_status	text

VALIDATION RULE

Produce mock outputs before unmasking. Preserve a clean rerun by an independent checker.

USE IN INTERPRETATION

The supplied script only illustrates fixture validation and descriptive statistics. A future locked analysis implements the trial estimand.

> **REVIEW RECORD**
>
> Record the schema version and its owner. Explain any deviations from these proposed fields in the released codebook and replication notes.

C11 · Public release register

DATA FIELDS

artifact_id	text
license	text
source_url	link
rights_review	categorical
released_version	text

VALIDATION RULE

Evaluate restrictions per artifact and distinguish protected from synthetic examples.

USE IN INTERPRETATION

Redistribution of a simulator asset can have different conditions from redistribution of code. Check both before packaging.

> **REVIEW RECORD**
>
> Record the schema version and its owner. Explain any deviations from these proposed fields in the released codebook and replication notes.

C12 · External replication checklist

DATA FIELDS

site	pseudonymous text
task_equivalence	categorical
source_pack	text
software_revision	text
outcome_agreement	text

VALIDATION RULE

An independent reproduction should record material differences, failures and deviations as well as successful runs.

USE IN INTERPRETATION

Publish enough details for others to interpret whether differences reflect population, apparatus or implementation.

> **REVIEW RECORD**
>
> Record the schema version and its owner. Explain any deviations from these proposed fields in the released codebook and replication notes.

D01 · Rubric calibration set

FICTIONAL INPUT

A response names the current revision, explains why its scope matches the scenario, and requests independent confirmation of a missing field.

EXPECTED REVIEW

The first, second, and third criteria can be credited even when the last action differs from the reference wording. Two reviewers independently score and compare before adapting the key.

REPRODUCTION CHECK

Retain both scores, the reference used and the revised scoring note.

> **INDEPENDENT RUN**
>
> Have a second reviewer record their judgment before they compare it with the expected review. Preserve the reasons for any disagreement.

D02 · Disagreement about scope

FICTIONAL INPUT

A response correctly recognizes a document mismatch but gives no reason to believe the second document is authoritative.

EXPECTED REVIEW

Give recognition credit independent of source-quality credit. Ask whether a narrower interpretation still permits a bounded conclusion.

REPRODUCTION CHECK

Check whether the rubric needs additional wording or the item itself is ambiguous.

> **INDEPENDENT RUN**
>
> Have a second reviewer record their judgment before they compare it with the expected review. Preserve the reasons for any disagreement.

D03 · Conflict-of-interest dry run

FICTIONAL INPUT

A product staff member asks a field coordinator for individual learner outcomes before analysis is complete.

EXPECTED REVIEW

Route the request to the written governance plan. Allow a technical troubleshooting summary only when it can be shared without identifying participants or exposing arms.

REPRODUCTION CHECK

Log the request, authorized response, any exception and decision owner.

> **INDEPENDENT RUN**
>
> Have a second reviewer record their judgment before they compare it with the expected review. Preserve the reasons for any disagreement.

D04 · Masking challenge

FICTIONAL INPUT

A learner’s answer mentions the tutor interface even though the outcome scoring file omits assigned arm.

EXPECTED REVIEW

Apply the predeclared masking rule consistently. Preserve the original and document the scope or uncertainty of unmasking.

REPRODUCTION CHECK

Report an assessor exposure count and an appropriate sensitivity check.

> **INDEPENDENT RUN**
>
> Have a second reviewer record their judgment before they compare it with the expected review. Preserve the reasons for any disagreement.

D05 · Contamination challenge

FICTIONAL INPUT

A protected assessment scenario duplicates the structure and record fragments of a practice exercise.

EXPECTED REVIEW

Review similarity between the actual evidence sets. Freeze an independent replacement before recruitment or apply the prespecified exclusion.

REPRODUCTION CHECK

Preserve the lineage record and report the effect on form equivalence.

> **INDEPENDENT RUN**
>
> Have a second reviewer record their judgment before they compare it with the expected review. Preserve the reasons for any disagreement.

D06 · Interrupted simulation

FICTIONAL INPUT

A sandbox session loses its connection before any answer is entered.

EXPECTED REVIEW

Code an apparatus interruption and follow the predetermined reschedule policy. Keep it distinct from a missed visit or an incorrect answer.

REPRODUCTION CHECK

Check whether downtime differs across conditions and document departures.

> **INDEPENDENT RUN**
>
> Have a second reviewer record their judgment before they compare it with the expected review. Preserve the reasons for any disagreement.

D07 · Hand-off verification

FICTIONAL INPUT

A credible tutor response says it cannot resolve a supplied conflict and names a source to verify.

EXPECTED REVIEW

Verify that the recommended source can settle the disputed point. Reward conditional explanation when relevant and document when referral remains unwarranted.

REPRODUCTION CHECK

Record outcome separately from substantive correctness for answerable items.

> **INDEPENDENT RUN**
>
> Have a second reviewer record their judgment before they compare it with the expected review. Preserve the reasons for any disagreement.

D08 · False reassurance check

FICTIONAL INPUT

A generated response cites a genuine document but its cited section does not support the claim.

EXPECTED REVIEW

Mark the traceability criterion independently of fluency. Classify severity by how much that mistake could change the learner’s interpretation.

REPRODUCTION CHECK

Log a source passage and precise failure type. Check comparators using the same rule.

> **INDEPENDENT RUN**
>
> Have a second reviewer record their judgment before they compare it with the expected review. Preserve the reasons for any disagreement.

D09 · Unavailable software case

FICTIONAL INPUT

A private workflow cannot be installed at a replication site.

EXPECTED REVIEW

Use a neutral sandbox representation and record the changed interface. Demonstrate form comparability before pooling it with any original workflow results.

REPRODUCTION CHECK

Report scope, version and portability limitations explicitly.

> **INDEPENDENT RUN**
>
> Have a second reviewer record their judgment before they compare it with the expected review. Preserve the reasons for any disagreement.

D10 · Privacy and release challenge

FICTIONAL INPUT

A transcript excerpt would reveal a rare combination of site and participant circumstances.

EXPECTED REVIEW

Replace the public example with a fully synthetic exemplar. Keep the restricted original in the controlled archive with reason and permissions.

REPRODUCTION CHECK

Check logs, export filenames, tables and prompts for indirect re-identification.

> **INDEPENDENT RUN**
>
> Have a second reviewer record their judgment before they compare it with the expected review. Preserve the reasons for any disagreement.

D11 · Sensitivity analysis rehearsal

FICTIONAL INPUT

A fictional missingness pattern is concentrated in one arm after the delayed visit.

EXPECTED REVIEW

Generate alternate outcome assumptions, calculate how the contrast moves and discuss how much unfavorable missingness would matter.

REPRODUCTION CHECK

This is a workflow example. Record assumptions and never present it as a participant estimate.

> **INDEPENDENT RUN**
>
> Have a second reviewer record their judgment before they compare it with the expected review. Preserve the reasons for any disagreement.

D12 · Replication readiness review

FICTIONAL INPUT

A new team receives a PDF, a versioned task bank and a synthetic scoring tool but no commercial product credentials.

EXPECTED REVIEW

Run the provided fixture checks and inspect dependency disclosures. Record which parts can be reproduced without access and what requires institutional agreement.

REPRODUCTION CHECK

Publish success and failure criteria separately from substantive research claims.

> **INDEPENDENT RUN**
>
> Have a second reviewer record their judgment before they compare it with the expected review. Preserve the reasons for any disagreement.