Nothing Left to Verify: How Simulation Design and AI Assistants Foreclose Information Literacy in Virtual Laboratories

Wang Haiying[1,2], Norhayati Hussin[2]*, Ezza Rafedziawati Kamal Rafedzi[3], Liao Meifang[4]

[1]School of Medical Information and Engineering, Ningxia Medical University, Yinchuan, China,

[2,3]Faculty of Information Science, Universiti Teknologi MARA (UiTM), Puncak Perdana, Selangor, Malaysia,

[4]Faculty of Information Science, Universiti Teknologi MARA (UiTM), Kelantan Branch, Malaysia,

***Corresponding Author**

DOI: https://doi.org/10.47772/IJRISS.2026.100800417

Received: 19 August 2026; Accepted: 24 August 2026; Published: 07 September 2026

ABSTRACT

Laboratory work exists to teach students how to look for information, weigh it, and decide what to trust. Virtual laboratories rarely give them the chance. This paper asks why, and sets out what an environment must retain if that chance is to survive. The starting point is simple: a virtual laboratory's information landscape is written by a developer, not encountered by a student. That makes its features design parameters rather than facts of the setting, and design parameters can be set differently. The paper is conceptual. It follows a theory adaptation design, using ideas from information literacy and information seeking research to put pressure on the inquiry-cycle account that currently guides virtual laboratory design. Literature was drawn purposively from physics education research, information science, and studies of generative AI in learning. University medical physics supplies the worked example. The analysis finds that virtual laboratories foreclose the information problem twice over. Scripted design strips out measurement variation, puts the accepted value on screen, and makes the interface self-sufficient; an embedded AI assistant then answers on request, and nothing obliges the student to check. Five informational attributes of the environment, plus one precondition governing their safe use, are placed within a four-layer framework, and seven propositions connect their settings to what students do and to what they learn. No data were analysed. The five attributes are proposed rather than validated, and the claim that low verification demand is the current default is stated as a premise awaiting documentary audit. The framework still yields workable design principles: treat measurement uncertainty as a variable, take the reference value off the screen, and make checking something that gets marked.

Keywords: virtual laboratory, information literacy, information seeking behaviour, generative AI, learning environment design

INTRODUCTION

Consider a medical student working through a virtual laboratory on X-ray attenuation. The accepted coefficient sits on screen. Her measurements come back identical on every run, and whenever she hesitates an embedded AI assistant supplies the next step. She finishes the exercise having met no gap to notice, no source to consult, no discrepancy to resolve. This is not an unusual laboratory. It is the ordinary one.

Two terms defined at the outset

An information problem, in the tradition running from Taylor (1968) through Belkin (1980) to Dervin (1998), is more than a blank waiting to be filled. Someone has to notice that something is missing. Someone has to work out where it might be found. And someone has to judge whether what turns up can be relied on. Informational foreclosure is the name used here for what happens when an environment removes one or more of those conditions before the student arrives, so that no information problem forms at all. The term says nothing about

scarcity. Information may be abundant; what has gone is the occasion to go looking for it, compare it and check it. Two mechanisms do this work, acting on different conditions, and the paper calls them the first and the second foreclosure. The term is unrelated to Marcia's (1966) identity foreclosure in developmental psychology.

What counts as a virtual laboratory

The framework covers computer-based environments where a student manipulates simulated apparatus and reads quantities off a model rather than off instruments: desktop simulations, three-dimensional laboratory environments, and hosted suites of experiments that stand in for a scheduled laboratory session. What unites them is that every number a student can obtain was written by somebody.

Two neighbouring cases fall outside. In a remote laboratory the student drives real apparatus from a distance, so the measurement stream still carries genuine variation; the first foreclosure does not apply there, though the second may. Screen-based analysis exercises built on a fixed dataset are excluded for a different reason. The evidence being weighed is not the student's own, so the attributes that govern how it was produced have nothing to describe.

The setting and the gap

Virtual laboratories are no longer a fringe option, and medical physics courses were among the first to adopt them. The case for adoption came largely from comparative studies showing that simulation can replace physical apparatus without measurable loss on conceptual understanding or procedural fluency (Finkelstein et al., 2005; de Jong et al., 2013). A second literature tells a less comfortable story. Comparing what experimental physicists do with what undergraduates do in a teaching laboratory, Wieman (2015) found little overlap. A multi-institution study found no measurable contribution of traditional laboratory instruction to physics content learning (Holmes et al., 2017). The diagnosis is structural. In the standard verification laboratory a student measures something, compares it with a number supplied in advance, and blames any difference on error. Almost nothing has to be decided. Put the decisions back and the picture changes: students repeatedly required to make quantitative comparisons and act on them showed real gains in critical thinking (Holmes et al., 2015). A laboratory course teaches little until the student is placed in a position of having to judge.

These two literatures have grown up with little contact. The gap has widened now that conversational agents built on large language models are being embedded directly in virtual laboratory software (Chen et al., 2026; Doğru & Faulconer, 2026).

The problem, the research questions and the contribution

Here is the problem. A well-built virtual laboratory can foreclose the very information problem that laboratory work exists to teach students to handle, and it can do so twice: first through scripted design, then through an embedded AI assistant. Because a developer writes the landscape rather than inheriting it, its informational features can be named as design parameters and set differently. That is the paper's central claim, and it is why what follows is a design framework rather than one more diagnosis.

Three questions organise the argument. What must an environment provide before an information problem can arise inside it? Which of those provisions does scripted design remove, and which does an AI assistant remove? And what testable propositions follow from the answers?

Medical physics supplies the worked example because a published attenuation coefficient carries unusually heavy institutional authority. Values of this kind come from bodies such as the International Commission on Radiological Protection (ICRP, 2007) and feed straight into dose calculation and patient safety. Deference to them is both especially likely and especially consequential, which makes the mechanism easy to see. Simulation-based training is well established in medical radiation science (Chau et al., 2022; Tene et al., 2024), and platforms such as the Virtual Environment for Radiotherapy Training are in routine classroom use (Jimenez et al., 2018), though published accounts of them centre on procedure rather than on measurement and checking.

LITERATURE REVIEW

What an information problem requires

Information science has long treated the information problem as a state that sets behaviour going, not as a quantity of missing text. Taylor (1968) distinguished four levels at which a need can exist and showed that moving between them is not automatic. Belkin (1980) described the initiating state as an anomaly in what the user knows. Dervin (1998) modelled it as a gap met while moving through a situation. Wilson (1999) placed these accounts inside a general model of need, seeking and use. Kuhlthau (1991, 2004) established the point that matters most here: the uncertainty accompanying that initiating state is a condition of seeking, not a fault to be designed away. All of this concerns what a person does once a need has arisen. The prior question, whether a setting lets a need arise at all, has attracted far less attention. An information problem is defined here as a situation in which students must recognise, locate and evaluate information in order to resolve uncertainty. Three conditions are required. An information gap exists when a student meets a difference between what is known and what the task requires (Belkin, 1980; Dervin, 1998). Uncertainty requiring resolution exists when the situation cannot be settled by reaching for a predetermined answer (Kuhlthau, 1991, 2004). Epistemic judgement demand, the demand that the student decide whether what has been found is relevant, credible and fit for the case at hand, exists when the answer cannot simply be taken as given (Rieh, 2002; Metzger, 2007; Association of College and Research Libraries [ACRL], 2015).

Foreclosing an information problem does not make information vanish. It removes one or more of these conditions in advance, so that the preconditions for seeking, evaluating and checking no longer hold. The definition also keeps the two mechanisms apart, because they act on different conditions. Scripted design works mainly on the gap and on uncertainty: it supplies reference values, strips out variation, and seals the edge of the environment. An embedded AI assistant works mainly on judgement demand, handing over answers that need not be appraised before use. One further property neither constitutes nor forecloses an information problem, but it decides whether the other three can be engaged with safely. Call it iteration reversibility: the cost of changing your mind or running the trial again. Its high pole means cheap, freely repeatable work; its low pole means one irrevocable run. Where a run cannot be undone, a student has no licence to be wrong, and every other condition is constrained by that fact (Kapur, 2008; Holmes et al., 2020).

What the design account lacks

It would be wrong to say that nobody has thought about withholding information. The assistance dilemma (Koedinger & Aleven, 2007), productive failure (Kapur, 2008) and work on scaffolding and its fading (Puntambekar & Hübscher, 2005) all treat giving or withholding as a decision with consequences. But these accounts are framed around what a student must build when no answer is handed over. They do not ask whether the student leaves the environment, consults a second source, sets two claims side by side, or works out that a supplied value does not fit the case. The difference is practical, not merely terminological. A task can withhold an answer completely and still prompt no information behaviour at all, if the student is meant to derive the answer rather than find it, and productive failure tasks are usually of exactly that kind. What these literatures supply is a language for building. What is needed here is a language for finding and appraising.

The inquiry cycle, which is what actually shapes virtual laboratory design (Pedaste et al., 2015), supplies neither. It names phases, includes orientation, conceptualisation, investigation, conclusion, discussion, and assumes throughout that students will get hold of whatever each phase needs. Whether the information is obtainable, from where, and at what cost in judgement is simply not a variable in the account. The consequence lands on the developer's desk. Working inside the inquiry cycle, there is no field in which to record the decision to display a reference value, no name for the state that results, and no way to weigh it against the usability that motivated it.

The increment over practice-based accounts of the information landscape

That a setting shapes which information practices are possible is not a new claim. Savolainen (2007) placed information behaviour inside social and contextual environments, and Lloyd (2010, 2017) took the position furthest, treating information literacy as a practice enacted within an information landscape whose features

decide which practices are available. The framework here sits inside that tradition. What it adds turns on where the landscape comes from.

Lloyd's landscapes are encountered. A workplace or a profession has an informational shape laid down over years of collective practice; nobody authored it, and the researcher's job is to describe it. A virtual laboratory is different. Every quantity a student can obtain, every source reachable without leaving the interface, every check the task demands, and the presence or absence of measurement variation are set by a developer before any student logs in. The landscape was written. And whoever wrote it can be asked to write it differently.

Two vocabularies for one behaviour

Three constructs do the analytical work. From the ACRL framework (2015) come three threshold concepts that bear directly on laboratory tasks. Authority Is Constructed and Contextual holds that a source's credibility depends on where it came from and how well it fits the question, not on some property it carries around with it. Research as Inquiry treats investigation as the successive sharpening of questions. Searching as Strategic Exploration treats looking as provisional and open to revision.

From Kuhlthau's Information Search Process come two more: the uncertainty principle, which makes uncertainty the trigger for seeking, and the zone of intervention, which makes the timing of help a design decision in its own right. Models of information problem solving break seeking into defining, searching, scanning, processing and organising (Brand-Gruwel et al., 2009), and a process broken into steps is a process that can be counted, which is what makes the behavioural layer of the framework observable rather than asserted. Recent restatements of information literacy for a world of generative systems ask what a student should now be able to do (Ali & Wilson, 2025). The question here is the mirror image: what an environment must keep if that capability is to be exercised at all. Table 1 sets the two vocabularies side by side.

Table 1: Construct correspondence between physics education research and information science

Laboratory phenomenon	PER construct	IS construct	Shared behavioural referent
Treating a published value as settled for the case at hand	Deference to authoritative values (Holmes et al., 2015; Wan, 2023)	Authority Is Constructed and Contextual (ACRL, 2015; Rieh, 2002; Lloyd, 2017)	Is the scope of an external claim treated as something to be established?
Judging whether a discrepancy is meaningful	Single-value versus distribution view of measurement (Buffler et al., 2001)	Evidence evaluation (Metzger, 2007)	Is evidence strength assessed before a conclusion is drawn?
Deciding what to do when data and model disagree	Experimental decision-making (Holmes et al., 2020)	Information problem solving (Brand-Gruwel et al., 2009)	Is an under-determined situation resolved by staged judgement?
Recognising that something is missing	Described (e.g., under question posing) but not operationalised as an act	Gap recognition (Belkin, 1980; Dervin, 1998)	Does the learner identify an information need?
Refining the question in light of results	Open-ended inquiry activity (Holmes et al., 2020; Pedaste et al., 2015)	Research as Inquiry (ACRL, 2015)	Is the question revised rather than executed?
Deciding where and when to look	Described but not operationalised as source selection	Searching as Strategic Exploration (ACRL, 2015; Kuhlthau, 1991)	Is seeking treated as contingent and revisable?

Note. PER = physics education research; IS = information science.

Two accounts already current in this territory do not predict the deference described in Section 4.3, and it is worth saying where each of them stops. Cognitive load theory (Sweller, 1988) explains neatly why scripted design is efficient, but it has nothing to say about the epistemic status of what gets supplied: taking the on-screen coefficient on trust and checking it against a second source look identical in load terms. Self-regulated learning models track how a student manages their own cognition (Panadero, 2017) and accommodate the drop in regulation reported under AI support (Fan et al., 2025). Yet a student may plan and monitor impeccably and still treat a published constant as settled, because accepting an authoritative value is not a failure of self-regulation. Authority Is Constructed and Contextual supplies the missing construct, and Lloyd (2017) supplies the reason it belongs as much to the setting as to the person. Epistemic cognition is not displaced by any of this. It underlies the difference between treating a measurement as a single true value and treating it as a distribution (Buffler et al., 2001; Wan, 2023), and it enters the framework as a moderator, as P7 states.

METHODOLOGY

The paper follows a theory adaptation design (Jaakkola, 2020), in which theory from one field is used to put pressure on theory established in another. The theory under pressure is the inquiry-cycle account of virtual laboratory design. The theory brought to bear comprises the ACRL Framework, Kuhlthau's Information Search Process, and models of information problem solving. The point of friction is precise. The inquiry cycle assumes students will look for information when they lack it, but it never treats the availability of that information as something the designer decides. Candidate attributes came from two places: the design features of the studies selected, and the public documentation of widely used virtual laboratory platforms. One was kept if its setting bears on information behaviour and can be observed from documentation or from interaction traces. Variation between implementations was recorded, not used to exclude. Five attributes vary. Iteration reversibility does not, sitting at or near its high pole almost everywhere, and is therefore treated as a precondition. Three candidates were dropped. Visual fidelity is not an informational property. Collaboration structure belongs to how students are grouped rather than to the informational environment, and returns in Section 6. Feedback immediacy was absorbed into verification demand, since what matters is not when feedback arrives but whether its arrival displaces an act of checking.

The Information Problem and Its Two Foreclosures

Five informational attributes of the environment

Whether the three conditions of Section 2.1 hold is a property of the environment, and it can be stated as a small set of variables a designer can set. Five are used throughout: D1 uncertainty availability, D2 ground-truth opacity, D3 evidential conflict, D4 source plurality, and D5 verification demand. Between them they describe what the environment settles in advance about finding, weighing and corroborating information, and nothing wider. Table 2 gives each one's definition, its two poles, and the work in each literature that grounds it. They can be set separately. They do not behave independently once set. D1 and D2 are generative: put both at their foreclosed pole and the rest have nothing to work on, because a display with no variation and a visible reference value leaves no discrepancy for D3 to register. D3 is what activates the others, since conflict is what makes a student notice a gap. D4 enables, by supplying the means of settling that conflict. D5 institutionalises, by deciding whether settling it counts for anything. The five are therefore not five independent switches. Shut the generative pair and D3 has no discrepancy to register, D4 no conflict to help settle, and D5 nothing whose settlement it could reward.

Table 2: Informational attributes of a virtual laboratory environment

Attribute	Definition	Poles (foreclosed → constituted)	Theoretical basis
D1 Uncertainty availability	Variation preventing a single observation from self-interpreting	No variation; identical repeated runs → stochastic, learner-configurable variation	Conceptions of measurement (Buffler et al., 2001; Wan, 2023); initiating uncertainty (Kuhlthau,

			1991)
D2 Ground-truth opacity	Whether the reference value is withheld from the learner	Accepted value displayed on screen → value withheld, dispersed or contested	Authority Is Constructed and Contextual (ACRL, 2015); cognitive authority (Rieh, 2002); deference (Holmes et al., 2015; Lee & See, 2004)
D3 Evidential conflict	Evidence that cannot be reconciled without judgement	No conflict; all evidence consistent → data conflicts with model, dataset or peers	Uncertainty and search initiation (Kuhlthau, 1991, 2004)
D4 Source plurality	Whether the task requires crossing the environment's boundary	Interface informationally self-sufficient → external sources must be consulted	Information problem solving (Brand-Gruwel et al., 2009); Strategic Exploration (ACRL, 2015); external verification (Visani Scozzi et al., 2026)
D5 Verification demand	Whether supplied information, including AI output, must be checked before use; a composite of provenance exposure	Answers on request, no check required → verification instrumented and assessed	Credibility assessment (Rieh, 2002; Metzger, 2007); cognitive forcing functions (Buçinca et al., 2021); verification in search (Mayerhofer et al., 2025)

The first foreclosure: scripted design

Virtual laboratories inherited the structure of the verification laboratory and, left alone, sharpened it. In medical physics the X-ray attenuation simulation usually has three features. Repeated runs at the same phantom thickness return the same number, because no detector noise was ever programmed in. The true coefficient is not an unknown to be estimated but a parameter written into the model, shown on screen as the standard value. And the interface hands over every quantity the task needs, with feedback that is instant and complete.

The first feature is where a simulation parts company with a bench. In physical apparatus variation is free and cannot be switched off. Noise, misalignment and thermal drift turn up whether the designer wants them or not, so even the most closed-down physical laboratory keeps D1 above zero. In a simulation that same variation has to be written in, and an unwritten parameter defaults to zero. What the bench supplies for nothing, the simulation must choose to supply. The first foreclosure is therefore a property of the medium as much as an inheritance from a teaching tradition, and recent work adding stochastic variation to widely used simulations treats exactly this parameter as a design choice (Liu et al., 2026).

Deference: the phenomenon both literatures describe

Early in a course, when students met a difference between their own measurements and an authoritative model, they characteristically doubted the measurement rather than the model (Holmes et al., 2015). The medical physics version is easy to picture. A student measures a coefficient that does not match the standard on screen and assumes the measurement is wrong, without ever asking whether the published constant was meant for this phantom material, this beam energy, this detector geometry. The finding has proved durable. Students draw conclusions from means without properly allowing for uncertainty (Wan, 2023), and an instrument administered across twenty institutions found that students remained weak at comparing measurements that carry uncertainty, even after laboratory instruction (Geschwind et al., 2024).

Research on trust in automated systems describes the same behaviour from another direction. Appropriate reliance means calibrating trust to what a system has been shown to do well, within a stated context of use, rather than extending it wholesale (Lee & See, 2004); and people will prefer algorithmic judgement to human judgement even where the algorithm's fit to the case has never been established (Logg et al., 2019). Deferring to

a number on screen and deferring to a fluent machine-generated answer are the same failure aimed at two different authorities. That is why the second foreclosure compounds the first instead of merely following it.

The pattern shows up again where the authority is a language model. Studying academic researchers working with large language models, Visani Scozzi et al. (2026) separate checking that a cited item exists and says what it claims from checking whether its claim ought to count, and report the second as the kind of checking least often done.

One qualification matters in a medical setting. The disposition at issue is not scepticism toward established standards. In clinical practice, deferring to values published by bodies such as the ICRP (2007) is correct professional behaviour. Authority Is Constructed and Contextual asks a much narrower question: does this published value apply to this case, at this material, this energy, this geometry? The question has a determinate answer, found by reading the source's own statement of the conditions under which it applies, and it is exactly the judgement a working medical physicist is paid to make. The target throughout is the appraisal of scope, not the contesting of authority, and the design principles in Section 6 should be read that way.

The second foreclosure: assistance without verification

Assistance carries the meaning it has in research on instructional support, where the question is never whether to help but how much to give and when. Hand over information and the student loses the chance to generate it; withhold it and the student may flounder unproductively (Koedinger & Aleven, 2007). Scaffolding that is never faded stops being scaffolding (Puntambekar & Hübscher, 2005). Kuhlthau's (2004) zone of intervention puts useful help where the student cannot proceed alone but could proceed with a nudge. On all three accounts, timing matters as much as content. A conversational agent that is always there and answers on demand takes timing out of the designer's hands entirely.

Verification carries the narrower meaning developed in credibility research: establishing whether a claim can be relied on before using it, by judging the source and the message (Rieh, 2002; Metzger, 2007). The framework splits what this one word has been made to carry. Required verification is a feature of the environment, and it is what D5 measures: whether provenance is visible at the interface, whether the task obliges a check, and whether any check gets marked. Spontaneous verification and transferred verification are features of the student's behaviour: checking done where the task did not ask for it, and checking that survives into a later task with no such requirement. The split matters. A proposition linking D5 to required verification would be true by definition; the propositions in Section 5 link D5 to the other two.

The two foreclosures differ in timing. The first is baked in before the student arrives; the second operates while the student works. In an unassisted environment, a student who lacks a value has to notice the gap, decide where to look, retrieve something, and judge whether it can be trusted. An assistant collapses all four into one exchange: a gap is stated, an answer comes back. What vanishes in the collapse is not the search. It is the verification. Offloading, as Risko and Gilbert (2016) put it, is adaptive or maladaptive depending on precisely what has been offloaded.

The supporting evidence is accumulating. AI assistance cut mental effort but also cut reasoning depth compared with conventional search, and the drop in germane load carried the whole of that effect (Stadler et al., 2024). People using conversational search queried more narrowly and more selectively (Sharma et al., 2024). Deliberately raising the cost of accepting an answer, through cognitive forcing functions, reduced overreliance, which shows that verification demand is a design variable rather than a fixed trait of the learner (Buçinca et al., 2021). In an introductory biology course, students preferred a language model's answer to a graduate teaching assistant's in forty per cent of cases and picked out machine-generated responses only forty-five per cent of the time, although the assistant beat the model overall on accuracy and effectiveness (Doğru & Faulconer, 2026). That study concerns replacing an assistant rather than a virtual laboratory in the sense of Section 1.2, and is used here only for the narrower point: once assistance is embedded, provenance is hard to recover.

Put an AI assistant into an environment whose reference values are already visible, whose data carry no variation, and whose tasks admit one defensible answer, and it does not restore the information problem. It removes what

was left of it. The student is now supplied not only with the answer to the measurement, but with the answer to what the measurement means.

The Framework And Its Propositions

The framework has four layers, and Figure 1 sets them out. Instructional design decisions (L1) — task openness, scaffolds, feedback, assessment criteria, agent configuration — set the five informational attributes (L2), subject to the iteration-reversibility precondition. Those attributes shape the seeking and evaluation a student actually does (L3). And that behaviour is the process through which information literacy develops, represented at L4 by the three ACRL threshold concepts named in Section 2.4. Three learner characteristics moderate the path from L2 to L3: beliefs about what a measurement is, prior knowledge of the domain, and metacognitive regulation. The contribution lies in L2. The other three layers are assembled from existing work.

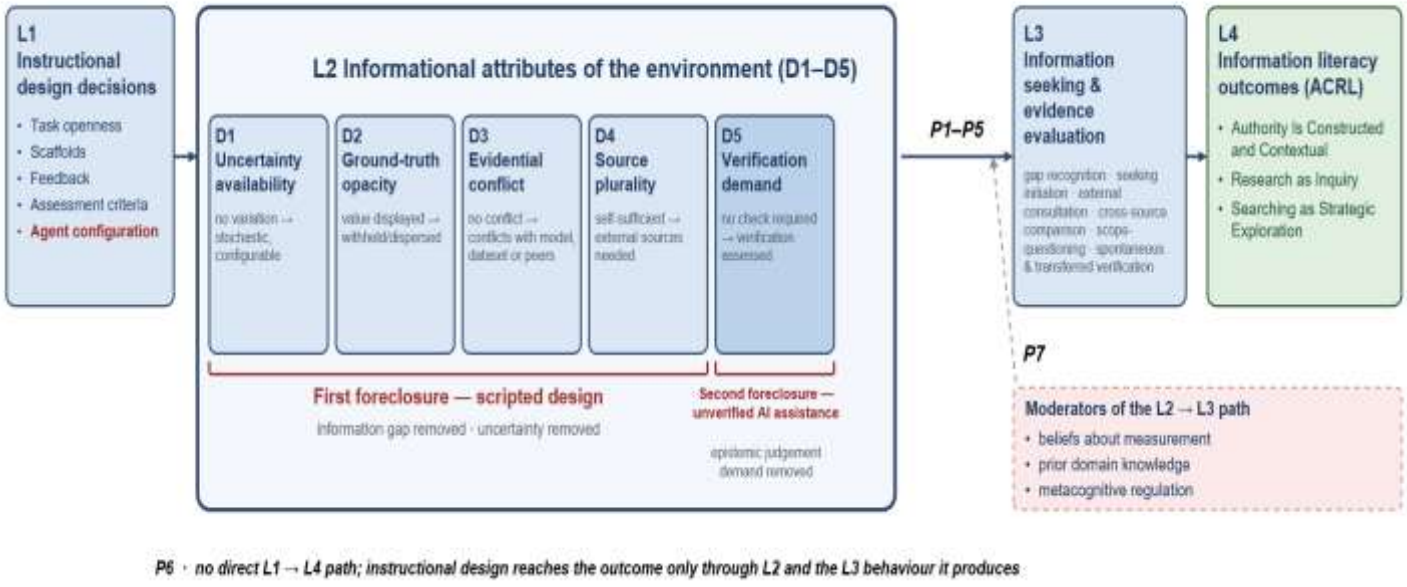

Figure 1: The informational foreclosure framework

Note. L1 assembled from the inquiry-cycle and scaffolding literatures (Pedaste et al., 2015; Puntambekar & Hübscher, 2005); L2 developed in this paper; L3 from models of information seeking and information problem solving (Kuhlthau, 1991, 2004; Brand-Gruwel et al., 2009); L4 from the ACRL Framework (2015). The brackets beneath the L2 band identify where each foreclosure acts and the condition it removes.

Scripted design acts on D1 through D4. It removes variation, displays the reference value, eliminates conflict, and makes the interface self-sufficient. D5 has little to bite on in an unassisted environment of this kind: when the interface supplies everything, there is not much left to demand a check on. An unverified AI assistant acts on D5 alone, but once D1 to D4 have been closed down, D5 is the last attribute through which the environment could still require a student to judge. Neither mechanism touches iteration reversibility. Nothing is worth repeating where there is no variation to characterise and no judgement to revise, though its high pole remains a background condition for everything P1 to P5 predict.

One piece of bookkeeping shapes the propositions. Deploying an agent is an L1 decision; D5 is the L2 attribute through which that decision reaches the student. P5 therefore describes an interaction within L2, not one across levels. The three-arm design recommended below — no assistant; an assistant with D5 high; an assistant with D5 low — sets agent availability and verification demand together for the same reason. The framework predicts that the first two arms will fall inside a prespecified equivalence margin on L3 behaviour, established by equivalence testing rather than read off a non-significant difference, and that the third will differ from both.

The central prediction follows from that. Adding a conversational assistant to an already scripted laboratory is

neither neutral nor a partial fix: it takes away the last condition under which information literacy could have been exercised. Two clarifications, beyond the split made in Section 4.4, keep the prediction falsifiable.

First, low verification demand is a configuration built around an assistant, not a property of assistants; an agent deployed with provenance visible and corroboration marked sits at D5's high pole, where the prediction simply does not apply. Second, the consequent is a specific dissociation, not a general claim that learning suffers. Spontaneous and transferred verification decline while task completion and reported satisfaction do not. Find the two moving together and the prediction is disconfirmed.

How common the foreclosed default actually is remains a premise of the framework rather than a finding of it. Settling it calls for a documentary audit: existing platforms coded against D1 to D5 from public documentation and hands-on use.

That audit is the first study this framework asks for. Because all three learner characteristics moderate the L2-to-L3 path, no proposition is stated as an unconditional main effect. Table 3 gives the seven propositions, what could be observed to test each, and what would count against it.

Table 3. The seven propositions, their candidate observations and their disconfirming findings

Proposition	Statement	What would disconfirm it
P1 (D1)	Uncertainty availability predicts gap recognition and seeking initiation, because one reading cannot interpret itself once the numbers move. How large the effect is depends on what the student takes a measurement to be (P7).	Gap recognition does not track D1; exposed variation is dismissed as noise rather than acted upon.
P2 (D2)	Ground-truth opacity predicts external consultation and cross-source comparison, since withholding the displayed value removes the cost asymmetry favouring an on-screen authority over a retrieved one. Magnitude is conditional on prior domain knowledge.	Withholding produces hesitation or guessing rather than consultation.
P3 (D3)	Evidential conflict predicts scope-questioning of an authoritative claim only where adjudication is both possible (D4) and affordable (iteration reversibility at its high pole). Absent D4 the response is guessing; where reversal is costly it is capitulation.	Scope-questioning occurs at equal rates with and without D4, or is unaffected by the cost of reversal.
P4 (D4)	Source plurality predicts greater search breadth but not greater evaluative depth where evidential conflict is absent and D5 remains at its foreclosed pole. Without a discrepancy to resolve, plurality supplies more places to look and no reason to weigh what is found. The proposition rests on the conjecture that search cost and judgement cost are separable, which it also puts at risk.	Breadth and depth move together, indicating that the two costs are not separable.
P5 (D5)	Where a conversational agent is available, the setting of verification demand determines the sign of the agent's effect on spontaneous and transferred verification, since offloading is adaptive or maladaptive according to what is offloaded (Risko & Gilbert, 2016). The claim is an interaction within L2.	The no-assistant and high-D5 arms fall outside the equivalence margin; or the low-D5 arm does not differ from both.
P6 (structure)	The effect of instructional design on information literacy outcomes runs only through attribute configuration and the L3	A direct L1-to-L4 path survives when L3 behaviour is

	behaviour it produces, with no direct path from L1 to L4.	controlled.
P7 (structure)	What a student takes a measurement to be moderates the path from D1 to L3, since responding to exposed variation with real inquiry requires seeing a measurement as a distribution.	Students holding the two views of measurement respond to exposed variation in same way.

P5 carries the framework's central practical claim and is where empirical work should begin. Iteration reversibility carries no proposition of its own and enters as a condition on P3. Prior knowledge and metacognitive regulation should be measured in any test of P1 to P5, but no directional prediction is made for them here.

Design Implications And Boundary Conditions

Because the attributes are configurable, the design principles follow directly, and a single X-ray attenuation laboratory shows all of them at work. Expose per-trial variation, so that one reading no longer interprets itself (D1); the PhET suite has already made statistical noise a controllable parameter in a purpose-built simulation and paired it with a data analysis platform (Liu et al., 2026). Take the standard value off the screen, so that the published coefficient has to be retrieved and its stated conditions of application checked (D2). Introduce a second dataset or a peer measurement that the model does not reconcile, and allow outside consultation before an answer is submitted, so that conflict arrives together with the means of resolving it (D3 with D4). Where an assistant is available, mark the corroboration performed rather than the number returned (D5).

All four principles rest on the iteration-reversibility precondition. A fifth concerns sequence rather than configuration: since P7 holds that beliefs about measurement moderate the response to exposed variation, those beliefs should be diagnosed before the variation is introduced (Buffler et al., 2001). Every principle here aims at the appraisal of scope described in Section 4.3, not at cultivating suspicion of published standards. The mechanism is not confined to medical physics. A chemistry laboratory chatbot that raised learner satisfaction without significantly improving learning shows the anticipated pattern (Chen et al., 2026), though that study was small, set in a secondary school class, and compared the assistant against unrestricted web search rather than against no assistance. What changes across settings is which reference standard is being deferred to, not the informational structure.

Four conditions mark where the framework stops. The evidence being weighed must be the student's own, since D1 and D2 have nothing to refer to where all evidence is supplied. Correctness must be adjudicable for D5 to mean anything, which genuinely open-ended research tasks do not satisfy. The unit of analysis is the individual, whereas laboratory work is mostly collaborative, and how students are grouped plausibly moderates every path in the framework. Disciplinary reach is illustrated rather than established: the attributes were derived from quantitative-measurement tasks with a checkable numerical ground truth, so extending them to judgement-based domains such as clinical image interpretation would mean rethinking D1 and D2 from the start.

Beyond these boundaries the main limitation is plain. No data were analysed, and the five attributes are proposed rather than validated. A coding exercise that kept forcing coders to invent a sixth category would show the set to be insufficient, and that operational test is offered in place of any claim to completeness. Two studies come first: the documentary audit described in Section 5, which tests the framework's premise, and the three-arm comparison under P5, which tests its central prediction.

CONCLUSION

The absence of information problems in virtual laboratories is not an accident of student ability. It is a consequence of how the informational conditions were designed, and this paper has given that state a name: informational foreclosure. Taking the three opening questions in turn, an information problem requires a gap, unresolved uncertainty, and the demand that the student judge; scripted design removes the first two, an embedded AI assistant removes the third; and seven propositions connect the settings of five informational attributes to what students do and to what they learn. The two mechanisms are therefore cumulative, not alternative. Because a virtual laboratory's information landscape is written rather than inherited, its features can

be named as design parameters and set otherwise. Nothing here says that virtual laboratories or conversational agents are bad. The claim is narrower and more demanding: uncertainty, evidence evaluation and verification are parts of inquiry that survive only if somebody decides to keep them. A developer who displays a reference value, suppresses variation, or embeds an unchecked assistant is making a decision about information literacy, whether or not it is recorded as one.

REFERENCES

1. Ali, R., & Wilson, M. L. (2025). Information literacy in the artificial intelligence era: A proposed framework. Communications in Information Literacy, 19(2), 242–270. https://doi.org/10.15760/comminfolit.2025.19.2.6
2. Association of College and Research Libraries. (2015). Framework for information literacy for higher education. American Library Association.
3. Belkin, N. J. (1980). Anomalous states of knowledge as a basis for information retrieval. Canadian Journal of Information Science, 5, 133–143.
4. Brand-Gruwel, S., Wopereis, I., & Walraven, A. (2009). A descriptive model of information problem solving while using internet. Computers & Education, 53(4), 1207–1217. https://doi.org/10.1016/j.compedu.2009.06.004
5. Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW1), Article 188.https://doi.org/10.1145/3449287
6. Buffler, A., Allie, S., Lubben, F., & Campbell, B. (2001). The development of first year physics students' ideas about measurement in terms of point and set paradigms. International Journal of Science Education, 23(11), 1137–1156. https://doi.org/10.1080/09500690110039567
7. Chau, M., Arruzza, E., & Johnson, N. (2022). Simulation-based education for medical radiation students: A scoping review. Journal of Medical Radiation Sciences, 69(3), 367–381. https://doi.org/10.1002/jmrs.572
8. Chen, C. M., Chen, B. J., & Huang, C. L. (2026). An Enhanced Al Chatbot to Facilitate Learning Performance in a Virtual 3D Chemistry Laboratory. Journal of Science Education and Technology, 1-22. https://doi.org/10.1007/s10956-026-10319-3
9. de Jong, T., Linn, M. C., & Zacharia, Z. C. (2013). Physical and virtual laboratories in science and engineering education. Science, 340(6130), 305–308. https://doi.org/10.1126/science.1230579
10. Dervin, B. (1998). Sense-making theory and practice: An overview of user interests in knowledge seeking and use. Journal of Knowledge Management, 2(2), 36–46. https://doi.org/10.1108/13673279810249369
11. Doğru, M. S., & Faulconer, E. K. (2026). ChatGPT as a virtual laboratory teaching assistant in undergraduate biology. Research in Science Education, 56(1), 379–399. https://doi.org/10.1007/s11165-025-10271-z
12. Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. British Journal of Educational Technology, 56(2), 489–530. https://doi.org/10.1111/bjet.13544
13. Finkelstein, N. D., Adams, W. K., Keller, C. J., Kohl, P. B., Perkins, K. K., Podolefsky, N. S., Reid, S., & LeMaster, R. (2005). When learning about the real world is better done virtually. Physical Review Special Topics: Physics Education Research, 1(1), Article 010103. https://doi.org/10.1103/PhysRevSTPER.1.010103
14. Geschwind, G., Vignal, M., Caballero, M. D., & Lewandowski, H. J. (2024). Using a research-based assessment instrument to explore undergraduate students' proficiencies around measurement uncertainty in physics lab contexts. Physical Review Physics Education Research, 20(2), Article 020105. https://doi.org/10.1103/PhysRevPhysEducRes.20.020105
15. Holmes, N. G., Keep, B., & Wieman, C. E. (2020). Developing scientific decision making by structuring and supporting student agency. Physical Review Physics Education Research, 16(1), Article 010109. https://doi.org/10.1103/PhysRevPhysEducRes.16.010109

16. Holmes, N. G., Olsen, J., Thomas, J. L., & Wieman, C. E. (2017). Value added or misattributed? A multi-institution study on the educational benefit of labs for reinforcing physics content. Physical Review Physics Education Research, 13(1), Article 010129. https://doi.org/10.1103/PhysRevPhysEducRes.13.010129
17. Holmes, N. G., Wieman, C. E., & Bonn, D. A. (2015). Teaching critical thinking. Proceedings of the National Academy of Sciences, 112(36), 11199–11204. https://doi.org/10.1073/pnas.1505329112
18. International Commission on Radiological Protection. (2007). The 2007 recommendations of the International Commission on Radiological Protection (ICRP Publication 103). Annals of the ICRP, 37(2–4).
19. Jaakkola, E. (2020). Designing conceptual articles: Four approaches. AMS Review, 10(1), 18–26. https://doi.org/10.1007/s13162-020-00161-0
20. Jimenez, Y. A., Thwaites, D. I., Juneja, P., & Lewis, S. J. (2018). Interprofessional education: Evaluation of a radiation therapy and medical physics student simulation workshop. Journal of Medical Radiation Sciences, 65(2), 106–113. https://doi.org/10.1002/jmrs.256
21. Kapur, M. (2008). Productive failure. Cognition and Instruction, 26(3), 379–424. https://doi.org/10.1080/07370000802212669
22. Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. Educational Psychology Review, 19(3), 239–264.https://doi.org/10.1007/s10648-007-9049-0
23. Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user's perspective. Journal of the American Society for Information Science, 42(5), 361–371.
24. Kuhlthau, C. C. (2004). Seeking meaning: A process approach to library and information services (2nd ed.). Libraries Unlimited.
25. Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. Human Factors, 46(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392
26. Liu, Q., Blackman, M., Geschwind, G., Carter, C., Perkins, K. K., & Lewandowski, H. J. (2026). Integrating noise into PhET simulations to promote student learning of measurement uncertainty. European Journal of Physics, 47(2), Article 025714. https://doi.org/10.1088/1361-6404/ae5774
27. Lloyd, A. (2010). Information literacy landscapes: Information literacy in education, workplace and everyday contexts. Chandos.
28. Lloyd, A. (2017). Information literacy and literacies of information: A mid-range theory and model. Journal of Information Literacy, 11(1), 91–105.https://doi.org/10.11645/11.1.2185
29. Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. Organizational Behavior and Human Decision Processes, 151, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005
30. Marcia, J. E. (1966). Development and validation of ego-identity status. Journal of Personality and Social Psychology, 3(5), 551–558. https://doi.org/10.1037/h0023281
31. Mayerhofer, K., Capra, R., & Elsweiler, D. (2025). Blending queries and conversations: Understanding trust, verification, and system choice in search and chat interactions. In Proceedings of the 2025 ACM SIGIR Conference on Human Information Interaction and Retrieval (pp. 168–178). ACM. https://doi.org/10.1145/3698204.3716454
32. Metzger, M. J. (2007). Making sense of credibility on the web: Models for evaluating online information and recommendations for future research. Journal of the American Society for Information Science and Technology, 58(13), 2078–2091.https://doi.org/10.1002/asi.20672
33. Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. Frontiers in Psychology, 8, Article 422. https://doi.org/10.3389/fpsyg.2017.00422
34. Pedaste, M., Mäeots, M., Siiman, L. A., de Jong, T., van Riesen, S. A. N., Kamp, E. T., Manoli, C. C., Zacharia, Z. C., & Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. Educational Research Review, 14, 47–61. https://doi.org/10.1016/j.edurev.2015.02.003
35. Puntambekar, S., & Hübscher, R. (2005). Tools for scaffolding students in a complex learning environment: What have we gained and what have we missed? Educational Psychologist, 40(1), 1–12. https://doi.org/10.1207/s15326985ep4001_1
36. Rieh, S. Y. (2002). Judgment of information quality and cognitive authority in the web. Journal of the American Society for Information Science and Technology, 53(2), 145–161.https://doi.org/10.1002/asi.10017

37. Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. Trends in Cognitive Sciences, 20(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002
38. Savolainen, R. (2007). Information behavior and information practice: Reviewing the umbrella concepts of information-seeking studies. The Library Quarterly, 77(2), 109–132. https://doi.org/10.1086/517840
39. Sharma, N., Liao, Q. V., & Xiao, Z. (2024). Generative echo chamber? Effect of LLM-powered search systems on diverse information seeking. In Proceedings of the CHI Conference on Human Factors in Computing Systems (Article 1033). ACM. https://doi.org/10.1145/3613904.3642459
40. Stadler, M., Bannert, M., & Sailer, M. (2024). Cognitive ease at a cost: LLMs reduce mental effort but compromise depth in student scientific inquiry. Computers in Human Behavior, 160, Article 108386. https://doi.org/10.1016/j.chb.2024.108386
41. Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4
42. Taylor, R. S. (1968). Question-negotiation and information seeking in libraries. College & Research Libraries, 29(3), 178–194. https://doi.org/10.5860/crl_29_03_178.
43. Tene, T., Bonilla García, N., Coello-Fiallos, D., Borja, M., & Vacacela Gomez, C. (2024). A systematic review of immersive educational technologies in medical physics and radiation physics. Frontiers in Medicine, 11, Article 1384799. https://doi.org/10.3389/fmed.2024.1384799
44. Visani Scozzi, M., Makri, S., & Madhyastha, P. (2026). Although powerful, it's not infallible: Investigating academic researchers' verification challenges with LLMs. In Proceedings of the 2026 Conference on Human Information Interaction and Retrieval (pp. 73–83). ACM. https://doi.org/10.1145/3786304.3787865
45. Wan, T. (2023). Investigating student reasoning about measurement uncertainty and ability to draw conclusions from measurement data in inquiry-based university physics labs. International Journal of Science Education, 45(3), 223–243. https://doi.org/10.1080/09500693.2022.2156824
46. Wieman, C. (2015). Comparative cognitive task analyses of experimental science and instructional laboratory courses. The Physics Teacher, 53(6), 349–351. https://doi.org/10.1119/1.4928349
47. Wilson, T. D. (1999). Models in information behaviour research. Journal of Documentation, 55(3), 249–270. https://doi.org/10.1108/EUM0000000007145