Schema Before Output: Metacognitive Schema-First Scaffolding in Computing Higher Education in the GenAI Era

Vjeran Bušelić
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
vbuselic@foi.hr

Dinko Horvat
Zagreb University of Applied Sciences
Vrbik 8, 10000 Zagreb, Croatia
dinko.horvat@tvz.hr

***Abstract*[1].** *Generative AI can produce the visible outputs of computing competence — architectural designs, database schemas, technical documentation, working code — before the schema those outputs are assumed to represent has formed. This paper reframes that as pedagogical design rather than detection or integrity. As a theoretical contribution, it proposes metacognitive schema-first scaffolding rather than testing it: schema formation should precede output as evidence of competence; AI should be a temporary, fading scaffold; and metacognition should be taught explicitly through the protocol hold–rethink–continue. Five teaching and assessment practices follow. Empirical validation remains future work, beginning with a Croatian institutional study.*

Keywords. computing education, generative AI, metacognition, schema-first learning, scaffolding, scaffold fading, hold–rethink–continue, competence formation, AI-assisted learning

1 Introduction

Generative AI did not enter education through a planning cycle, a procurement decision, or a curriculum committee. It entered through the everyday infrastructure already used by students and educators: phones, browsers, and natural-language interfaces that were operational before institutional policy could respond. By the time universities began discussing frameworks, students were already using the tools.

This is not simply an institutional failure. Education systems change - curriculum cycles, accreditation processes, and assessment frameworks are slow by design. Their mission is to produce reliable and comparable competence credentials for the large interconnected system, not one particular domain. The real problem is that the surrounding environment has changed faster than those structures can adapt. UNESCO acknowledged in 2023 that publicly available generative AI tools were emerging faster than regulatory frameworks could adjust (UNESCO, 2023). The OECD Digital Education Outlook 2026 confirms that even knowledge about integrating generative AI into teacher-designed learning scenarios remains insufficient (OECD, 2026).

The practical consequence is a pedagogical paradigm shift. The visible output of learning — architectural designs, database schemas, technical documentation and working code — can now be generated by AI before the learner has necessarily performed the cognitive work through which competence forms. The standard assessment principle — this student produced this output, therefore; this student has this competence — has become unreliable. This creates a formation gap: a hidden difference between the output a student can produce with AI support and the schema-based competence that output is assumed to represent.

The problem is not primarily that students may be dishonest, but that AI tools can now produce exactly the kind of output that assessment was designed to certify.

Most students in higher education encounter GenAI as a disruption to their learning practice. A medical student using GenAI as a new tool in a field where core professional acts — diagnosis, clinical judgement, and patient interaction — remain formally human-led and institutionally regulated. The work is changing, and many AI-assisted tools must be learned, but the core professional role has not been redefined. An IT student faces a different situation: the tool that is restructuring their learning environment is also restructuring the professional field they are entering at an unprecedented speed.

[1] This paper is one of a pair developed in parallel for this conference. It addresses the institutional and pedagogical response to formation-stage AI adoption in IT education. A companion paper addresses the risk as experienced by the IT student. Each paper stands independently.

Their target is moving while they are being trained to hit it. This double exposure — GenAI reshaping both how they learn and what they will be expected to do (adapting O'Brien and Leichenko's [2000] double-exposure framework to a single technology) — is becoming the norm in IT education and has few parallels in other disciplines.

Professional pressure is already evident. Stack Overflow's Developer Survey reports widespread adoption of AI tools among developers, particularly for code generation, debugging, documentation, and testing – professional skills current IT students are developing right now (Stack Overflow, 2025). The World Economic Forum's Future of Jobs Report 2025 identifies information and technology services as sectors facing significant digital transformation and increased pressure on routine programming roles (World Economic Forum, 2025). This trend is not just a forecast for IT students; it reflects the current reality of the professional environment they are about to enter. IT students are joining the field during a period of rapid transition.

The lack of transitional memory makes this problem significant. While IT educators have experienced platform changes and developed new schemas gradually over time, IT students lack this experience. They have no practical way to distinguish when AI assistance supports a competency and when it replaces it. For these students, the logical choice is to use the available tools, improve results, and meet requirements. The cognitive cost is not immediate, but delayed, accumulating unnoticed during education and emerging later when independent judgment is required, even if AI remains available. This paper defines this moment as the point of delayed recognition: when students realize that although a tool can help with performance, it cannot replace a schema they never developed.

For computing educators, this clarifies both the urgency and the opportunity. Their role is not to discourage students from using essential tools, but to provide timely instruction that integrates AI use during schema formation. This paper supports educators navigating the gap between institutional delays and student adoption. While it does not present a comprehensive institutional solution, it introduces a metacognitive schema-first scaffolding framework for immediate action. The core principle is hold–rethink–continue, a metacognitive control loop that positions GenAI as a scaffold for learning rather than a shortcut.

2 From Productivity to Scaffolded Learning

AI offers real productivity benefits in education - the presented evidence highlights the associated formation costs.

2.1 When better output does not mean better learning

Bastani et al. (2025) present direct empirical evidence of formation risk. In a preregistered field trial in high school mathematics with nearly 1,000 students, unrestricted ChatGPT-4 access improved AI-assisted practice performance but reduced subsequent independent assessment performance. The study also finds that students were often unaware of this decline because their confidence was shaped by their AI-supported practice. Fan et al. (2025) describe the related habit as metacognitive laziness: when AI handles monitoring, evaluation, and checking, learners may underuse those capacities themselves. Shen and Tamkin (2026) provide convergent evidence in a developer-learning context, where AI-supported task completion did not translate into stronger post-task mastery in conceptual understanding, code reading, and debugging. Bassner et al. (2025) add a programming-specific finding: AI support reduced stress and improved exercise scores without automatically improving code understanding.

These studies highlight a structural risk: AI may enhance the appearance of competence without ensuring the underlying schema is developed. A schema is a mental model representing a student's internal understanding of why something works, not just how to perform it. Its value is in transferability. Once formed, students can explain, adapt, and apply knowledge in new contexts. If AI generates results before this structure is established, students may produce outputs without genuine understanding.

Even when students recognize this risk, the underlying incentives often prevail. Caosun and Aral (2026) model skill dynamics under AI adoption and demonstrate that fully informed students may still rationally choose to use AI heavily. Immediate benefits, such as higher assignment scores and project deliverables, can outweigh the future cost of skill erosion. This is not a result of naivety, but a rational response to conflicting performance and capability signals. When incentives drive behaviour, awareness-based guidance is insufficient.

Caosun and Aral's distinction between “performance-extracting” deployments, in which AI operates independently of student judgment, and “skill-preserving” deployments, in which students must configure, direct, or verify AI contributions, offers a useful framework for the practices proposed in Section 5.

2.2 Metacognition: the mechanism that is missing

The question this raises is not whether AI should be used, but how a student is supposed to notice when it has crossed from scaffold to substitute. Scaffold means temporary learning support that helps the student progress while responsibility gradually returns to the

learner. GenAI can play this role only if the student remains cognitively engaged with the underlying schema rather than accepting the output as a finished product. Metacognition, defined as the capacity to monitor, evaluate, and regulate one's own thinking process (Flavell, 1979), is a well-established concept. It refers to "thinking about thinking," an internal dialogue in which educators and students actively plan, monitor, and assess their learning, regardless of whether they explicitly use the term. In most learning environments, metacognition develops gradually through practice and challenge. For instance, it is practical students' strategy for passing the test, by understanding their own knowledge and abilities and the nature of the test (written, oral, …).

In GenAI environments, metacognition is no longer only a habit that forms in the background — it becomes a prerequisite for recognizing when AI has replaced independent thinking rather than supported it. The metacognitive question — did I learn this, or did the tool do it for me? — does not arise reliably on its own. It must be deliberately trained.

The theoretical framework for this habit is already established. Zimmerman (2002) identified three cyclical phases of effective self-regulation: (1) forethought, which includes goal-setting, planning, and anticipating outcomes; (2) performance, involving monitoring, self-instruction, and attention management; and (3) self-reflection, which covers evaluation, attribution, and adaptive adjustment. Panadero (2017), in his widely cited review of six Self-Regulated Learning models, confirms these phases are consistent across theoretical traditions.

GenAI disrupts each phase: forethought is bypassed when answers are immediately available, performance monitoring is replaced by accepting outputs, and self-reflection is skipped when learners accept AI evaluations instead of their own.

The **hold–rethink–continue** protocol introduced in Section 4.3 puts this cycle into practice. Educators are responsible for designing conditions that restore this process.

2.3 The educator's design obligation

By design, schooling rewards answers much more than the inspection of how answers are produced. Even in math, where the process of arriving at the answers is often rewarding, students won't get credit unless the results are right, too.

GenAI intensifies this pattern: the tool delivers fluent answers immediately, with no hesitation, definitely without the invitation to examine the procedure. The metacognitive habit of checking the procedure, not just the answer, is what would alert a student to the gap between a solution and the reasoning behind it. It does not develop on its own, which is precisely why educators must introduce it explicitly. Bassner et al. (2025) introduce what they call the comfort trap: students do not naturally choose the most effective learning interaction. In their research, students rated the unrestricted AI tool as more helpful in every area, even though it did not improve learning and a guided option was available. Students tend to prefer interactions that require less mental effort, even if those interactions are less beneficial. As a result, educators cannot expect students to develop strong metacognitive habits on their own when using GenAI. These habits need to be built into the way tasks and assessments are designed, since the students most likely to use AI passively are also the least likely to notice the risks involved.

The underlying challenge is not new. Over a decade ago, OECD (2015) tracked educational technology investments in eighteen countries and found no improvement in reading or mathematics; outcomes stagnated or declined in high-investment nations. The issue was not the technology itself, but the order of implementation: tools were introduced before pedagogical practices were established. Kirschner, Sweller, and Clark (2006) explain that minimally guided instruction consistently underperforms compared to direct instruction for novices, as they lack the ability to distinguish relevant from irrelevant information. Uncontrolled AI access in computing education creates a similar situation: the environment is highly responsive, guidance is not sequenced, and learners without prior schemas cannot differentiate productive engagement from plausible-sounding output. Educators must redesign tasks and assessments to address this recurring gap that each new wave of technology reopens.

Educators are in a different position. The OECD Teaching and Learning International Survey (TALIS) 2024 provides representative, comparative information on how lower secondary education teachers use AI. Educators are already using AI primarily for preparation and productivity tasks: on average, 68% of those who use AI, use it to efficiently learn about and summarise topics they teach, and 64% use it to generate lesson plan. By contrast, non-use is largely explained by insufficient skills, lack of confidence, limited training opportunities, or uncertainty about pedagogical value. (OECD, 2026, pp. 18-19).

That productivity is not automatically a pedagogical gain. In a pre-registered field experiment in K–12 schools, Sungu et al. (2026) found that giving teachers a generative-AI support tool — used mainly for the preparation tasks above — reduced students' intrinsic motivation and, among lower-performing teachers, their measured performance: a principal–agent gap in which the teacher's time saving did not translate into a student benefit. That study is in schools rather than higher education, so what carries over is the mechanism, not the effect size — but the substitution logic this paper traces for students applies to educators' own use just as well: AI adopted for output rather than engagement can undercut the outcome it appears to serve.

This highlights the need to position educators as key mediators. Their proximity to classroom practice enables them to redesign tasks, clarify requirements for explanations, and update assessment rubrics. Their direct experience introducing AI in practice provides a personal reference for understanding how to build a schema amid uncertainty. As a result, they are well-placed to introduce metacognitive habits before IT students reach the point of delayed recognition. This paper proposes the hold–rethink–continue cycle as a teachable metacognitive control loop, detailed in Section 4.3. Its aim is to re-engage students with AI through understanding rather than dependency.

3 Methodology: Theoretical Research Design

This paper is theoretical research in computing education. It does not present new empirical data and does not assert that metacognitive schema-first scaffolding has been validated as a comprehensive instructional intervention. The schema-first ordering is not merely hypothetical: in direct response to generative AI, one of the authors redesigned a first-year informatics Information Literacy and Critical Thinking course — a prerequisite aimed at independent reasoning — so that students build their own mental model of a problem before using AI for prompt formulation, and defend it in a live debate rather than a written essay, keeping the assessed output resistant to substitution (Bušelić, 2023). This is a design precedent, not evidence of effect: it shows the ordering is teachable, and it motivates the empirical study proposed in Section 6.

The objective is to develop an original conceptual framework for computing education in the GenAI era by synthesizing established theories of competence formation, scaffolding, tacit knowledge, metacognition, and relational understanding with recent empirical findings on AI-assisted learning. The complementary student-facing perspective on this risk is developed in a companion paper presented in parallel.

The theoretical framework of this paper is built upon six complementary foundations. First, the CC2020 Knowledge–Skills–Dispositions model defines the competencies that computing educational programs should develop and certify (ACM/IEEE-CS/AAAI CC2020 Task Force, 2020). Second, Polanyi's (1966) tacit knowledge theory shows that professional competence involves embodied understanding developed through practice, which cannot be fully conveyed by explicit demonstration. Third, scaffolding theory provides the pedagogical basis for schema-before-output instruction. Vygotsky (1978) establishes the zone of proximal development, in which support must fade as learner capacity grows. Wood et al. (1976) define scaffolding as temporary support for tasks beyond unaided capability, and Reiser (2004) explains how scaffolds structure the task while preserving its thinking challenge. Vendrell and Johnston (2026) extend this to LLM-supported critical thinking in higher education. Fourth, Flavell's (1979) metacognition framework and Schraw and Dennison's (1994) metacognitive awareness model provide the basis for treating metacognition as a teachable control mechanism in GenAI-rich learning environments. Fifth, Skemp's (1976) distinction between instrumental and relational understanding clarifies why correct performance may conceal the absence of transferable understanding. Sixth, established learning science provides the normative substrate for the teaching practices in Section 5: deliberate practice (Ericsson, Krampe & Tesch-Römer, 1993), self-regulated learning (Zimmerman, 2002), retrieval practice (Dunlosky et al., 2013), and feedback (Hattie & Timperley, 2007). Together these identify the conditions competence formation requires — conditions the framework proposes to reinstate.

The paper's original theoretical contribution is the framework of metacognitive schema-first scaffolding (MSFS), defined and developed in Section 4. In brief: schema formation must precede output as evidence of competence; AI must function as a fading scaffold rather than a substitute; and the two are held together by the taught metacognitive loop hold–rethink–continue. The framework draws on metacognitive monitoring and regulation (Flavell, 1979; Schraw & Dennison, 1994; Zimmerman, 2002), scaffolding as temporary support (Wood et al., 1976), and the distinction between surface performance and relational understanding (Skemp, 1976).

The sources for this paper are intentionally divided into two distinct categories, and the selection does not represent a comprehensive literature review. It serves a different purpose.

The first category includes established theoretical foundations: widely cited, empirically supported models whose core claims are uncontested in their fields. Examples include scaffolding and the zone of proximal development (Wood et al., 1976; Vygotsky, 1978), deliberate practice (Ericsson et al., 1993), self-regulated learning (Zimmerman, 2002; Panadero, 2017), retrieval practice (Dunlosky et al., 2013), and feedback (Hattie & Timperley, 2007). These works establish the normative basis for competence formation.

The second category includes recent empirical studies from GenAI-era educational research, such as Bastani et al. (2025), Fan et al. (2025), Shen and Tamkin (2026), Bassner et al. (2025), and Yan et al. (2025). These studies do not serve as proof of the proposed framework. Instead, they provide supporting evidence that the framework is relevant to real, observable situations in current practice. Those two categories are complementary: established learning science defines the requirements for competence formation, while recent GenAI studies demonstrate

that current practice is systematically disrupting these conditions.

As this paper is theoretical, its contribution is supported through analytical rather than experimental evidence. Three forms of confirmation are provided: First, theoretical coherence: the framework is based on the established theoretical foundations outlined above. Second, triangulation conceptual framework: different theoretical traditions identify the same educational requirement, that learners must develop transferable structures of understanding rather than simply produce correct outputs. Third, empirical plausibility: the studies reviewed in Section 2.1 indicate that output performance and independent competence formation may diverge, and that scaffolded or constrained AI assistance may better support learning than unrestricted answer production.

4 The Framework: Metacognitive Schema-First Scaffolding

The main idea is simple: in computing education, a student should build understanding before their work is taken as proof of skill. More precisely, metacognitive schema-first scaffolding (MSFS) — the paper's central contribution — is a conceptual framework: a lens for diagnosing where AI-mediated formation fails and for designing against it, not a validated instructional method or assessment model. It integrates three established principles into one response to AI-mediated formation risk, and each principle names one part of the term:

- Schema-first — schema formation, not output, is the primary instructional goal; a task is valued for the mental model it builds, not the artefact it yields. Once AI can generate the output, that output no longer evidences the understanding behind it, so that understanding must become the explicit target (Section 4.1).
- Scaffolding — AI is positioned as temporary, calibrated support that fades as the schema forms, never as a substitute for the cognitive work through which it forms (Section 4.2).
- Metacognitive — the framework is made teachable through hold–rethink–continue, the loop by which a student monitors their understanding, evaluates what the tool produced, and directs AI toward schema-building rather than answer-production (Section 4.3).

The contribution is the integration, not any single principle: none works alone. Schema-first without scaffolding is an instruction without a mechanism; scaffolding without a schema goal optimises better output rather than understanding; metacognition without either activates anxiety rather than learning. MSFS uses scaffolding to create the conditions under which metacognition can activate, and metacognition to keep the scaffolded work aimed at schema formation rather than production — which is why assessment must then certify the schema, not the output AI can now supply (Section 4.4). Section 5 offers concrete teaching and assessment practices that illustrate how the framework can be applied in a course. They are examples, not validated methods — and not the framework itself.

4.1 Schema before output

The clearest example comes from mathematics. A child learning addition may start with apples, blocks, or fingers, which make the operation visible. However, the educational goal extends beyond the apple calculation. It is the schema of addition: a transferable mental structure that later applies to any countable object, then to symbolic algebra, then to abstract structures the student has never seen. Once the schema is formed, any representation works. Without it, the student can only handle the specific cases they have been shown. Skemp (1976) named this distinction instrumental understanding — knowing the rule and producing the correct result — versus relational understanding — knowing why the rule works and being able to adapt it. The difference is invisible in the output. A student with relational understanding and a student with only instrumental understanding can produce identical answers. The gap appears the moment the problem changes.

GenAI makes this distinction urgent in computing education. A student can ask AI to generate working code, a correct database schema, or a well-structured requirements document. The output is often indistinguishable from what a competent student might produce independently. But if the student accepts the output without forming the underlying schema — without understanding why this structure, why this decomposition, why this trade-off — then the output is not yet evidence of independent competence. The next task, the changed constraint, and the novel problem cannot be addressed independently by accepting the previous output again.

They require the schema. If it was never formed, the tools are present but the understanding is not; this is where the formation gap becomes visible.

The educational question is therefore not whether students can produce required outputs with AI. The question is whether they can explain the reasoning behind what was produced, transfer it to a new case, diagnose its failure, and adapt it when the context changes.

Schemas are not built through successful performance. They are sharpened by error — by the moment a solution breaks, a constraint changes, or an approach that seemed correct reveals its limits. That friction is not a failure of instruction. It is the mechanism that exposes the untrained parts of a schema and forces them to develop. A student who never encounters this friction does not experience smooth learning; they experience a hidden ceiling: it

stops exactly where the difficulty was removed. When AI removes this friction before the schema is complete, it may remove the sharpening mechanism along with the obstacle.

4.2 Scaffolding rather than substitution

AI can support schema formation or bypass it. The difference is not in the tool — it is in how the tool is used. When a student asks AI to explain why a recursive function terminates, compare two sorting algorithms and justify the choice for a specific case, or generate test cases after predicting edge cases themselves — AI is functioning as a scaffold. It extends what the student can do while keeping the student cognitively accountable for the schema. When a student asks AI for a finished solution and submits it without engaging the underlying logic — AI is functioning as a substitute. The schema that the task was designed to form does not form.

There is a simple test for students: after using AI, can you explain what was created, why it was organized that way, and what would change if the situation were different? If you can, the AI helped guide your learning. If you cannot, the tool just replaced your own work. This is not about right or wrong. It is about growth. If a student cannot answer these questions, they have not yet learned what the task was meant to teach.

Reiser (2004) explains that effective scaffolds operate through two mechanisms: structuring, which reduces unnecessary complexity, allowing learners to focus on essential elements, and problematizing, which maintains the visibility and challenge of intellectually demanding components. A scaffold that eliminates all difficulty ceases to function as one and instead becomes a bypass. Artificial intelligence occupies this boundary, and its role depends on both the educator's task design and the student's engagement with the tool.

Wood, Bruner, and Ross's foundational study points out a key requirement that clarifies this boundary. Effective scaffolding needs two models: a theory of the task and a theory of the learner (Wood et al., 1976, pp. 97–98). The first is about knowing what a good solution looks like and how to get there, which general-purpose AI can often simulate. The second is about understanding where the learner is, what they already know, where their understanding is incomplete, and what kind of support will help them grow without skipping over their current abilities. General-purpose prompt-based AI tools do not reliably offer this unless they are part of a carefully designed educational approach.

A general-purpose AI system responds uniformly to prompts and cannot calibrate its support to the learner's developmental state without teacher-designed diagnostic structure. Without a theory of the learner, there is no reliable basis for calibration; without calibration, neither structuring nor problematizing can be targeted precisely to the developmental gap. This distinction also aligns with recent GenAI-specific scaffolding literature. Vendrell and Johnston (2026) argue that LLMs should be integrated into higher education through design principles that preserve critical thinking rather than outsource it to the tool.

AI does not bypass scaffolding by design. However, when it is used primarily as an answer-production system rather than as part of a scaffolded learning design, it may bypass the developmental function that scaffolding is meant to serve. A scaffold is also, by definition, temporary. The goal is not permanent support but gradual withdrawal — fading the scaffold as the learner's own schema grows strong enough to carry the task independently. In computing education, this means that AI support should change form as competence increases: from hints and guided comparisons, to explanation and transfer, to independent performance checks.

The aim is to ensure that answer-substitution fades as schema ownership grows.

The fading requirement comes from Vygotsky's (1978) idea of the zone of proximal development. Support should decrease as the learner becomes more capable; otherwise, the learner never takes over control. AI cannot meet this need because it always responds to every question with the same level of detail, no matter what the learner already knows. Fading must be built into the task itself by deciding when AI can be used, when it should be limited, and when learners must work independently to show they can complete the task on their own.

4.3 Hold–rethink–continue as a student protocol

Hold–rethink–continue was introduced earlier as a metacognitive control loop. In this section, it is presented as a student-facing protocol that can be explicitly taught and deliberately practiced.

Hold is a short pause before accepting AI output — enough to ask: do I really understand what this is, and why should I accept it?

Rethink is the reconstruction of the schema behind the output. Ask yourself: Why does this solution work? What assumptions are behind it? How would things change if the input was different, the constraints were stricter, or the scale was bigger? What could go wrong, and how would you notice? This is not just a checklist. It is about making a habit of stepping back from the result and looking at your own understanding. This is metacognition in its simplest, most practical form.

Continue refers to re-engaging with AI using a more specific question. Rather than requesting an answer, the student might ask for a hint, a contrasting case, or an explanation of why an approach fails under a particular constraint. After reflecting, the student returns to the tool with new questions and a better ability to assess the response.

Consider as an illustration, presented in Figure 1., a student asked to design a schema recording which students take which courses and the grades they earn.

A first prompt to a generative AI tool — "*give me the SQL tables for this*" — returns three tables that create without error, including an enrolment table the student did not anticipate.

```
CREATE TABLE student (
    student_id    INT PRIMARY KEY,
    full_name     VARCHAR(100) NOT NULL,
    enrolled_year INT
);
CREATE TABLE course (
    course_id  VARCHAR(10) PRIMARY KEY,
    title      VARCHAR(100) NOT NULL,
    credits    INT NOT NULL
);
CREATE TABLE enrollment (                          -- the junction table
    student_id INT         NOT NULL REFERENCES student(student_id),
    course_id  VARCHAR(10) NOT NULL REFERENCES course(course_id),
    semester   VARCHAR(10) NOT NULL,
    grade      CHAR(1),
    PRIMARY KEY (student_id, course_id, semester)
);
```

Figure 1. Schema returned by a generative AI tool for the enrollment assignment: three tables, including the enrollment junction table

Hold. The student pauses before accepting. The code runs; that is not the question. The question is: I described students and courses — why are there three tables, and why is the grade in the third one?

Rethink. The student reconstructs the schema behind the output. Why can a grade not sit in student (a student earns many) or in course (a course awards many)? What does the key (`student_id, course_id, semester`) actually enforce — and what happens to a retaken course if semester is removed from it? What would break if all three tables were merged into one? Working these through surfaces the many-to-many relationship, the grade as a property of the pairing rather than either entity, and the update anomalies a single flat table would create.

Continue. The student returns to the tool with a sharper question — not "is this right?" but "show me a concrete case where putting the grade in the student table fails." The tool demonstrates the anomaly; the student now evaluates the answer from a formed position rather than accepting it from an empty one.

The protocol encourages students to use AI in ways that keep their own thinking and understanding active. Educators can help by assigning reflection annotations, asking for oral explanations, or including structured phases in tasks. The protocol is not idiosyncratic to this framework. Reviewing thirty-nine studies of generative AI and critical thinking, Nasr et al. (2026) independently derive a structured human-in-the-loop model in which the learner plans, assesses, explores, evaluates, refines and applies — arriving, from a different literature and without reference to this work, at the same conclusion: unstructured use degrades the capability, and structure is what preserves it. Hold–rethink–continue operates one layer above that sequence, as the decision to engage at all; hold fires before planning begins.

The goal for students is to internalise this metacognitive habit before the delayed recognition point arrives.

4.4 Assessment as evidence of schema formation

If we agree that the mere output is no longer sufficient evidence of competence, assessment must make the process of building it (schema formation) visible. The difficulty this creates is not new, and it did not begin with generative AI. In 1990, the APA Delphi panel of forty-six critical-thinking experts recorded that the readily assessable skill dimension should not be allowed to distort the conceptualization of what matters, and that self-regulation — the capacity to monitor and correct one's own reasoning — is "extremely difficult to assess using standard paper and pencil instruments" (Facione, 1990). The gap between what is assessable and what is important is therefore a standing condition of education, identified by the field's own consensus body more than three decades ago. Generative AI did not create it. It industrialised it: what was a known limitation of assessment has become the ordinary case. This does not mean abandoning submitted outputs. Instead, we should add explanations, examples of transfer, error analysis, and independent performance checks.

This also completes a distinction the paper's own foundation draws but has not yet used. CC2020 separates Knowledge and Skills from Dispositions, and is explicit that the third is not a softer version of the first two: a graduate who holds knowledge and skills without the dispositions has codified competence without professional judgment (ACM/IEEE-CS/AAAI CC2020 Task Force, 2020). Dispositions form through repeated practice and **cannot be demonstrated in an output** — no artefact shows that its author would have checked had no one been watching. Where AI substitutes for the cognitive work, it can supply the Knowledge and reproduce the Skills while leaving the Dispositions unformed, and no inspection of the product will reveal the difference.

This heuristic shortcut shows how schema-first scaffolding can be used in assessment design. The framework is based on ideas from the literature; it is not an explicit taxonomy from any single source.

Table 1. Assessment targets ordered by resistance to AI substitution

Assessment target	**What it asks the student to demonstrate**	**AI risk if assessed only as product**
Local production	Produce syntactically valid code, diagram, or document	**High** — AI generates reliably

Comprehension	Explain what an output does and trace its behavior	**Moderate** — oral or constrained explanation reveals gaps
Design rationale	Justify structure, decomposition, and trade-offs	**Moderate** — contextual justification requires student ownership
Error diagnosis	Find, explain, and correct failures in a flawed output	**Lower** — requires schema activation
Transfer	Apply the same schema to a new problem or changed constraint	**Lower** — exposes whether schema is independent of the original output
Metacognitive regulation	Describe what was attempted, where AI was used, and how the result was checked	**Lower** — when students must explain their process, AI use, and verification steps

Note. The framework described and compared through Table 1. is the authors' synthesis.

Risk levels serve as heuristics to show how easily students might meet assessment targets using AI, without demonstrating independent schema formation. **High risk** means AI can reliably generate responses, allowing students to meet the target without engaging their own understanding. **Moderate risk** indicates AI can assist, but contextual specificity limits direct substitution. **Lower risk** requires schema activation, so AI output cannot replace context-specific performance.

This pattern keeps showing up. When assessments focus on the lower rows, they are much harder for AI to replace. If a student cannot explain the reasoning behind their output structure, it shows a gap in their understanding that is not evident from the output alone.

5 Teaching and Assessment Practices

The proposed framework offers teaching practices that educators can use right away in their current courses, without waiting for a full curriculum redesign. These are not just extra improvements added to existing assessments. Instead, they challenge three main assumptions: that output alone shows competence, that difficulty should always be removed, and that AI use is simply a matter of allowing or banning it. Each practice directly addresses one of these ideas. These approaches also align with new GenAI design research, which views LLMs as tools to support critical thinking rather than as replacements for student judgment (Vendrell & Johnston, 2026).

Each of the proposed practices restores a condition that learning science, before the rise of GenAI, regarded as essential for building competence. Ericsson, Krampe and Tesch-Römer (1993) showed that expertise comes from deliberate practice, which means working on tasks that challenge current abilities, getting feedback on mistakes, and staying engaged with difficult work. When students have unlimited access to AI, these conditions often disappear. When educators see these practices as a way to restore important learning conditions, they shift from policing AI use to helping students build real-life skills.

5.1 Scaffold AI access through task phases

One practical improvement is to decide not just if AI is available for a task, but when it is available. Breaking a task into phases like independent attempt, AI-supported exploration, explanation, and transfer helps students build their own understanding while still letting AI help later on. In the first phase, students try the task themselves, sketch a design, or predict what a correct answer would look like. This gets them thinking about the problem and shows what they already know before AI steps in. In the next phase, students use AI for hints, comparisons, or explanations instead of asking for full solutions. Asking questions like "why does this work?" helps them learn more than just asking AI to do the work. Finally, students explain what was produced and apply the same logic to a new situation. These steps make it clear whether AI was used as a learning tool or merely as a shortcut for output.

The pre-GenAI phase has a specific name. Dunlosky et al. (2013) identified retrieval practice as one of the most robustly supported learning techniques. Its effectiveness comes not from reviewing correct information, but from attempting recall before the answer is available, which serves as the consolidation mechanism. AI inverts this sequence structurally: the answer appears before retrieval has been attempted. Requiring an independent attempt, design sketch, or prediction before AI access operationalises retrieval practice in a computing context.

5.2 Use explanation gates

An explanation gate is a point in the task where progress depends on the student's ability to explain the schema behind the output — not only what it does, but why it is structured as it is. It may be oral, written, or interactive. Its function is not to detect AI use. It is to make schema formation a precondition for completion. Ask yourself why this structure was chosen instead of another. What would happen if you removed one part? Where could this approach fail, and how might you notice problems early? You can’t answer these

questions just by repeating what AI says. You need to think through the ideas on your own.

5.3 Design error-analysis tasks

Error-analysis tasks are especially useful for computing educators in the GenAI era. In these tasks, students receive a flawed output, such as code with a logic error, a database schema with an integrity problem, or a system design with a structural bottleneck. They are then asked to find the error, explain why it is a problem, and suggest a fix. This approach brings back important elements that AI-driven coursework can sometimes remove too soon: real uncertainty, the challenge of diagnosing issues, and the moment when students realize the difference between what they thought they knew and what they actually understand. In an error-analysis task, the schema cannot go untested. It must be put to use, confront the flaw, and either stand up to the test or show where it falls short.

5.4 Use deliberate non-use as scaffold fading

Scaffold fading — the deliberate withdrawal of support to verify that the schema can carry the task independently — is the natural counterpart to scaffolded AI use.

Periodic tasks that require students to work without AI are not meant to punish or restrict AI use. Instead, they help confirm whether the scaffold's support has been effective. For students' integrity, it is important not to position these tasks as tests of their honesty. Rather, they serve as checkpoints where support is removed, not because the tool is flawed, but so the learner can see if they can complete the task on their own. Being able to perform independently under these conditions shows that a student has truly learned the material, rather than just producing the right answers. This is the best way for both students and educators to see if real learning has taken place. Yan et al. (2025) studied this directly. In their three-group randomized controlled trial, both a proactive and a passive GenAI agent used the same AI technology; the only difference was how students interacted with them. While support was available, all three groups performed similarly. When support was removed, only the proactive group kept improving, while the passive group returned to the same level as the group without AI. This shows that scaffold-fading is a real and measurable effect: giving a task with AI support, followed by one that requires explanation or transfer without AI, helps reveal whether students have truly learned or just relied on support.

When we look at this through the idea of deliberate practice, the discomfort of working without AI—the effort, uncertainty, and mistakes—is not just a side effect of learning. Instead, it is the main way learning happens. The conditions that Ericsson et al. (1993) say are needed for developing competence are exactly what working without AI brings back.

5.5 Track schema formation longitudinally

A single assessment provides only a necessary snapshot. A formation trajectory requires data across time. The signal that matters is not whether students use AI, but whether their independent performance converges with their assisted performance as instruction progresses. A student whose assisted and independent performance diverge across a semester is on a formation-risk trajectory. A student whose independent performance rises toward assisted performance provides evidence that the intended schema is forming. Low-stakes diagnostic assessments — not high-stakes credentialing instruments, but formation-tracking checkpoints — give both students and educators the information needed to identify the formation gap before the delayed recognition point becomes the first evidence that independent competence has not formed.

Longitudinal tracking requires feedback that informs the learner's next step, not only feedback that evaluates the current output. Hattie and Timperley (2007) distinguish four feedback levels: task (is the product correct?), process (what strategy produced it?), self-regulation (can the student monitor and adjust?), and self (what does this mean for the learner's development?). Task-level feedback — the default in product-based assessment — is the least effective for long-term competence formation; the trajectory tracking proposed here requires process and self-regulation level feedback. The self-regulation level has a practitioner precedent: in the course introduced in Section 3, students assessed their own competence before and after as a metacognitive exercise — building the self-monitoring habit, not measuring whether competence had formed (Bušelić & Rajković, 2024). The convergence signal itself still requires the independent, unassisted assessment described here. The structural implication from Caosun and Aral (2026) also applies here: under high-AI-use conditions, assisted performance is no longer a reliable baseline. Periodic unassisted assessment is therefore not punitive — it is the only measurement context in which the convergence or divergence signal becomes visible.

5.6 A worked example: one assignment across the practices

The following worked example is illustrative only. It makes the framework's application concrete and constitutes no empirical claim; consistent with Section 3, this paper remains theoretical.

The practices above are not five separate assignments but five ways of designing one. The enrolment schema of Section 4.3 illustrates the whole sequence presented in Figure 2.

Independent attempt (5.1). Before any AI, the student sketches an entity–relationship diagram by hand and states the relationship in words: a student takes many courses; a course is taken by many students. This is retrieval practice — it reveals whether the student recognises a many-to-many relationship at all. A student who stops at two boxes and a connecting line has exposed the exact gap that AI would otherwise fill invisibly.

AI-supported exploration (5.1). The student now uses AI not to produce the schema but to compare designs: contrast recording the grade on the enrollment row with holding it in a separate assessment table — when is each better? The student remains the one choosing.

Explanation gate (5.2). Progress depends on the student explaining why the relationship needs a third table, why the grade belongs to the pairing, and what the composite key enforces. Pasted output cannot answer this.

Error analysis (5.3). The student receives a flawed variant that still executes, and must find, explain, and correct it:

```
CREATE TABLE enrollment (
    student_id      INT         NOT NULL REFERENCES student(student_id),
    course_id       VARCHAR(10) NOT NULL REFERENCES course(course_id),
    semester        VARCHAR(10) NOT NULL,
    grade           CHAR(1),
    instructor_name VARCHAR(100),            -- (1)
    PRIMARY KEY (student_id, course_id)      -- (2)
);
```

Figure 2. A flawed variant of the enrollment table, used for the error-analysis task

Both faults are silent at creation time. (1) `instructor_name` depends on the `course`, not on the `student-course-semester` pairing — a transitive dependency that will duplicate and eventually contradict itself. (2) The key omits semester, so a retaken course collides with its own earlier attempt: the schema quietly encodes the rule *no student may ever retake a course*. Neither fault is visible from the fact that the code runs; only an activated schema catches them.

Transfer. The constraint changes: a student may retake a course, and both attempts must be kept. The correct schema survives; the flawed one fails. Whether the student can adapt it shows whether the schema is theirs or the tool's.

The same assignment scored against Table 1.:

Table 2. The enrollment assignment mapped onto the Table 1. assessment targets

Table 1. target	What the student is asked to do here	AI substitution risk
Local production	Write the CREATE TABLE statements	**High** — AI generates reliably
Comprehension	Explain what the enrollment table does	Moderate
Design rationale	Justify the composite key and the third table	Moderate
Error diagnosis	Find and fix the two faults above	Lower
Transfer	Adapt the schema to allow retakes	Lower
Metacognitive regulation	State where AI was used and how it was checked	Lower

As figure 2. indicates, the assignment does not change across these rows — only what is asked of it does, and that is what separates an assessment AI can satisfy from one it cannot.

6 Research Agenda

The framework proposed in this paper is conceptual: its value lies in providing a disciplined basis for designing and testing an intervention. The central empirical question is direct: does metacognitive schema-first scaffolding produce better independent competence than unrestricted AI use, AI prohibition, or product-only assessment? No existing study answers this for computing education as an integrated framework. Testing it would proceed in two phases.

The first would be a baseline. Before the framework can be introduced or measured, it would help to know where practice currently stands: how computing educators who already work with generative AI design their tasks and assessment, and how far they recognise the formation gap described here. A survey of such educators — for instance at the Faculty of Organization and Informatics, University of Zagreb (FOI), and Zagreb University of Applied Sciences (TVZ), subject to institutional support — would provide early data from Croatian institutions and add a Central European perspective to a literature still largely Anglo-American. Its aim would be modest: not to test the framework, but to map current practice and to prompt educators to examine their own assessment habits. It is intended as an impetus, not a precondition.

The second phase would test the framework itself. Several questions follow from the analysis:

- Does scaffolded AI access — phased tasks, explanation gates, and error-analysis requirements

— improve independent transfer performance compared with unrestricted AI access?
- Which scaffold types are most effective for schema formation in computing education: hints, worked contrasts, explanation requirements, delayed access, or error diagnosis?
- Does explicit metacognitive training — teaching hold–rethink–continue as a student protocol — reduce passive answer acceptance and improve independent performance?
- Does the divergence between AI-assisted and independent performance narrow across a semester when schema-first scaffolding is applied?
- Can educators reliably assess schema formation through explanation, transfer, and error-analysis tasks, and does this assessment predict independent professional performance better than product-only assessment?
- Can hold–rethink–continue be taught to students who do not yet possess the metacognitive capacity it presupposes — and if not, what educator-supplied structure substitutes for that capacity until it forms?

The last question is the framework's own hardest, and it is not rhetorical. Nasr et al. (2026) find the benefit of generative AI for critical thinking conditional on the metacognitive skills needed to interrogate it — precisely the skills the formation-stage student has not yet built. Every protocol of this kind, including the one proposed here, therefore presupposes some measure of the capability it is meant to form. Whether that circle can be broken by educator-supplied structure, or whether it marks a genuine limit, is an empirical question this framework cannot settle by argument.

Such a study would compare scaffolded and unscaffolded computing cohorts, using independent — not AI-assisted — performance as the outcome measure, and would track whether the assisted-to-independent gap narrows over a semester.

Whether and when this becomes possible would depend on institutional support. Our own institutions would be one place to begin, but the framework and the two-phase path outlined here are offered to any computing programme to take up — and to adapt to its own context, constraints, and questions. It is put forward now precisely so that such work, in this form or a modified one, can be designed against it.

7 Conclusion

Generative AI does not eliminate the necessity of computing education; rather, it clarifies its objectives. If AI can generate outputs that traditional assessments were intended to validate, educational frameworks must specify the knowledge, understanding, self-regulation, and transfer skills students require before such outputs can serve as evidence of competence. This need for clarification should not be viewed as a crisis but as an opportunity to establish a more robust model than the previous output-focused approach.

This paper makes three central arguments.

First, information technology students encounter a formation risk that is structurally distinct from the broader AI-in-education challenge. It is characterised by double exposure, the absence of transition memory, and a professional field already being transformed by the very adoption logic these students will soon face.

Second, the mechanism underlying both the emergence and mitigation of this risk is metacognition, defined as the ability to reflect on the learning process and assess whether genuine understanding has been achieved.

Third, metacognitive schema-first scaffolding offers educators a practical framework for immediate implementation.

This includes sequencing AI integration through schema development, maintaining productive friction, instructing students in the **hold–rethink–continue** protocol, and confirming schema independence through scaffold fading.

Educators need not wait for institutions to change or for new policies to be implemented. The integrated framework awaits empirical validation, but the principles it draws together — schema before output, scaffolding that fades, and explicit metacognition — are well established in learning science and can be applied now, with the tools educators already have, treating the development of students' understanding as no less important than the results they produce.

Computing educators occupy a central role in facilitating **schema formation**. The hold–rethink–continue principle is **concrete enough** to introduce in the first week of a course and **consequential enough to shape** students' future professional practices.

- **Hold** before accepting the assumption that output proves competence.
- **Rethink** what each assignment is designed to form.
- **Continue** — forward, with AI as a scaffold for learning rather than a shortcut around it.

References

ACM/IEEE-CS/AAAI CC2020 Task Force. (2020). Computing curricula 2020: Paradigms for global computing education. *ACM*. https://doi.org/10.1145/3467967

Bassner, P., Lenk-Ostendorf, B., Beinstingel, R., Wasner, T., & Krusche, S. (2025). Less stress, better scores, same learning: The dissociation of performance and learning in AI-supported programming education. *Computers and Education: Artificial Intelligence*, 100537.

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high

school mathematics. *Proceedings of the National Academy of Sciences*, 122(26), e2422633122.

Bušelić, V. (2023). Teaching information literacy and critical thinking skills in ChatGPT time. 2023 *International Conference on Computing, Networking, Telecommunications & Engineering Sciences Applications (CoNTESA)*, Zagreb, Croatia.

Bušelić, V., & Rajković, I. (2024). Teaching generic skills with/or against ChatGPT. *2024 47th MIPRO ICT and Electronics Convention (MIPRO)*, Opatija, Croatia.

Caosun, M., & Aral, S. (2026). The augmentation trap: AI productivity and the cost of cognitive offloading [arXiv preprint]. https://arxiv.org/abs/2604.03501

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques. *Psychological Science in the Public Interest*, 14 (1), 4–58.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100 (3), 363–406.

Facione, P. A. (1990). Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction (The Delphi Report). *American Philosophical Association*.

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.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 34(10), 906–911.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77 (1), 81–112.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work. *Educational Psychologist*, 41 (2), 75–86.

Nasr, N. R., Tu, C.-H., Sujo-Montes, L., Haniya, S., Yen, C.-J., & Hossain, S. (2026). Generative artificial intelligence impact on critical thinking in higher education: A comprehensive bibliometric analysis and systematic review. *Journal of Computers in Education*.

OECD. (2015). Students, computers and learning: Making the connection. PISA, *OECD Publishing*.

OECD. (2026). OECD Digital Education Outlook 2026: Exploring effective uses of generative AI in education. *OECD Publishing*. https://doi.org/10.1787/062a7394-en

O'Brien, K. L., & Leichenko, R. M. (2000). Double exposure: assessing the impacts of climate change within the context of economic globalization. *Global Environmental Change*, 10(3), 221–232.

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, 8, 422.

Polanyi, M. (1966). The tacit dimension. *Doubleday*.

Reiser, B. J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *The Journal of the Learning Sciences*, 13(3), 273–304.

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology*, 19(4), 460–475.

Shen, J. H., & Tamkin, A. (2026). How AI impacts skill formation *[arXiv preprint]*. https://arxiv.org/abs/2601.20245

Skemp, R. R. (1976). Relational understanding and instrumental understanding. *Mathematics Teaching*, 77, 20–26.

Stack Overflow. (2025). 2025 Stack Overflow Developer Survey. https://survey.stackoverflow.co/2025/

Sungu, A., Lira, B., & Duckworth, A. L. (2026). Generative AI can harm teaching *[SSRN working paper]*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=7007339

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

Vendrell, M., & Johnston, S.-K. (2026). Scaffolding critical thinking with generative AI: Design principles for integrating large language models in higher education. *Computers and Education: Artificial Intelligence*, 10, 100572.

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. *Harvard University Press.*

Yan, L., Martinez-Maldonado, R., Jin, Y., Echeverria, V., Milesi, M., Fan, J., Zhao, L., Alfredo, R., Li, X., & Gašević, D. (2025). The effects of generative AI agents and scaffolding on enhancing students' comprehension of visual learning analytics. *Computers & Education*, 234, 105322.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*, 41 (2), 64–70.