Classwork as Evidentiary Practice

ICPS, semantic structure, independent transfer, and the near-elimination of grading ambiguity

Peter Alexander Thought Museum Grading Architecture Series, Position Paper I Version 1, August 2026

Abstract. The word *classwork* commonly names an administrative category before it names a defensible form of evidence. It may absorb participation, compliance, notebooks, group contribution, effort, behavior, practice, and whatever else occurs during a class period. Such breadth makes the category convenient and epistemically unstable: a point may be awarded, yet no one can say exactly what the point represents. This paper reconstructs classwork as a disciplined evidentiary practice. It develops a mathematical adaptation of Myrna B. Shure's Interpersonal Cognitive Problem Solving tradition, retaining both the acronym ICPS and the learner-facing affirmation *I Can Problem Solve*. The adapted sequence moves from a fully modeled contextual problem, through dialogic interrogation and a systematically varied collaborative case, to an independent open-notes transfer artifact. Only the final artifact is scored. The paper coordinates this sequence with the phenomenological architecture of *From Taxonomy to Tact*, research on worked examples, scaffolding, self-explanation, analogical transfer, mathematical problem representation, multiple representations, and assessment validity. It proposes a fixed twenty-point analytic architecture–SARC–20: Semantic Representation, Analytical Accuracy, Reasoning Traceability, and Contextual Completion–together with task-specific criteria, criterion codes, explicit status rules, and a high-resolution evidence ledger. The objective is not to eliminate professional judgment, which would be neither possible nor desirable, but to eliminate unaccountable judgment. A classwork score should be a compressed, auditable claim about an observable mathematical performance produced under stated conditions.
Keywords: classwork; ICPS; I Can Problem Solve; mathematical problem solving; semantic structure; problem representation; transfer; worked examples; scoring rubrics; grading validity; pedagogical tact

"The old initiated; the new merely 'conditions'."
– *C. S. Lewis*, The Abolition of Man

1 The category before the evidence

The departmentally required grading architecture assigns twenty percent of the course grade to *Classwork*. The percentage is settled. The meaning is not. A category may be fixed administratively while remaining conceptually empty, and an empty category does not stay empty for long. It becomes a receptacle for whatever the teacher happens to observe: the visible notebook, the raised hand, the completed worksheet, the cooperative posture, the rapid answer, the good attitude, the quietness interpreted as disengagement, the talk interpreted as leadership, the work that appears careful, or the behavior that has tested patience. Each item may matter within teaching. It does not follow that each item is mathematical evidence or that all of them may be converted into the same kind of point.

This position paper begins from a severe requirement:

The governing claim

A classwork point is not a reward distributed during class. It is a compressed claim about the quality of an identifiable mathematical artifact produced under publicly specified conditions and judged by criteria established before the artifact was encountered.

The distinction is not verbal. It separates grading from the ordinary economy of classroom approval. A teacher may encourage a question, protect a hesitant student from premature exposure, correct a disruption, require materials, praise persistence, redirect attention, or insist upon professional conduct. Those acts belong to teaching and classroom governance. They do not become evidence of mathematical mastery merely because they occur in a mathematics room.

The problem is more acute when the population is heterogeneous. A student may speak often because the material is easy, because public fluency is part of the student's social identity, or because talk protects against sustained private work. Another may speak little because the material is difficult, because the problem is being considered carefully, because public error has acquired the meaning of humiliation, or because the student has already perceived a relation that others are still verbalizing. As Alexander (2026) argues, visible behavior underdetermines the lived instructional situation. To convert that behavior directly into points is not only imprecise. It confuses evidence about one construct with evidence about another.

A category that can mean everything finally means nothing. The purpose of this paper is therefore not to decorate the inherited word *classwork* with a longer definition. It is to identify the exact object that will be graded, the conditions under which it will be produced, the inferences that its score may support, the inferences it may not support, and the public procedures by which each point will be awarded or withheld.

1.1 What this document is, and is not

This is a foundational design document. It is not the final student-facing policy. The student-facing policy will eventually be a dense compression of the argument developed here. Its brevity will be warranted by a deeper structure that remains available for administrative review, parent inquiry, scoring calibration, and future revision.

The paper is also a position paper rather than a report of a completed experimental validation. The original ICPS tradition has an empirical and programmatic history in interpersonal and social problem solving. The mathematical adaptation proposed here is new. It is theoretically grounded, operationally specified, and designed for local study, but it should not borrow an empirical warrant that belongs to another domain. That boundary will be stated repeatedly because exact grading begins with exact claims.

Finally, this paper does not attempt to settle Assessment or Homework. It marks their borders only far enough to prevent category leakage. Assessment will require its own account of independent performance under higher-stakes conditions. Homework will require its own account of practice completed outside direct supervision. The present task is narrower and more exacting: to make the twenty percent called *Classwork* intelligible from the first assigned point to the final trimester average.

2 The position: classwork as a defined evidentiary object

The proposed category contains one and only one graded object: the culminating independent artifact of an ICPS–M cycle. The cycle is an instructional sequence. Its final artifact is the evidentiary object. The distinction may be represented as follows.

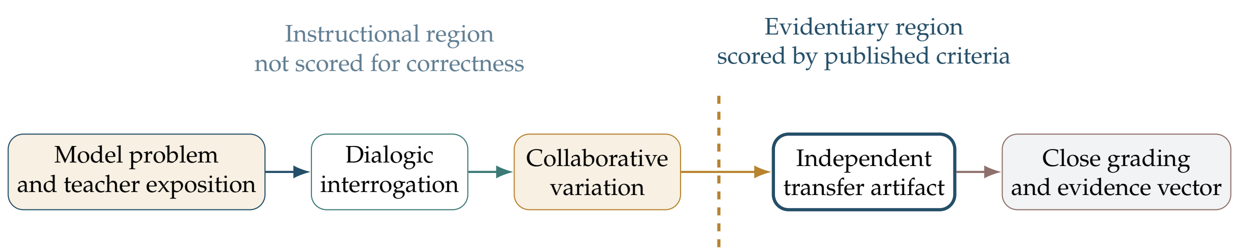

Figure 1: *The ICPS–M cycle separates the work of instruction from the object of grading. The boundary is deliberate: supported practice is not converted into points, while the final independent artifact is.*

The first region includes explanation, note-taking, questioning, comparison, provisional errors, collaborative reasoning, and teacher response. These are not lesser activities. They are the means by which the mathematical field becomes available. Their freedom from correctness grading is part of their instructional function. Students must be able to propose, revise, and expose misunderstanding without every provisional move becoming a permanent numerical trace.

The second region begins when the independent transfer task is issued under stated conditions. At that point the student may use the permitted notes and tools, but must independently represent the problem, select and execute a strategy, make reasoning visible enough to be audited, and return the result to the context. The submitted work is then scored by a task-specific instantiation of a stable analytic framework.

2.1 A formal definition

For purposes of the course, **Classwork** is defined as follows:

Operational definition

Classwork is the set of scored, independent, supervised mathematical artifacts produced at the conclusion of documented ICPS–M instructional cycles. Each artifact is completed under specified access conditions, scored on a fixed twenty-point architecture with task-specific criteria published in advance, and entered into the gradebook with a criterion vector and concise evidentiary annotation.

This definition excludes ordinary notes, copied examples, ungraded practice, oral participation, group contribution, behavior, preparedness, effort, and work completed outside the specified independent phase. Those matters may be observed, taught, required, praised, corrected, or documented elsewhere. They are not hidden components of the Classwork score.

2.2 Boundaries among the three grading categories

The categories can be distinguished by the conditions under which evidence is produced, not merely by the location in which an assignment happens.

Table 1: *Category boundaries within the required 70/20/10 architecture. Only the Classwork column is fully specified in this paper.*

	Assessment (70%)	**Classwork (20%)**	**Homework (10%)**
Primary function	Independent demonstration under quiz, test, midterm, or final conditions	Supported learning culminating in a supervised independent transfer artifact	Deliberate practice and preparation outside the immediate instructional period
Source of score	Assessment response	Final ICPS–M artifact only	Verified completion or quality under a separately defined homework policy
Support before scoring	Instruction precedes assessment but is not embedded in the event	Modeling, dialogue, and collaborative variation are integral to the cycle	Resources and assistance may be available; provenance is less controlled
Inference	What the student demonstrates under the specified assessment conditions	What the student can make of a documented instructional sequence in a calibrated transfer task	What practice was undertaken and, where warranted, what was produced outside direct supervision

The table prevents an apparently minor but consequential mistake: using the Classwork category to repair limitations elsewhere. Classwork will not become a participation grade because participation is difficult to classify. It will not become a homework-completion grade because outside work is difficult to verify. It will not become a miniature test category merely because the final artifact is independent. Its distinctiveness lies in the instructional sequence from which the artifact emerges.

3 Scores as claims: evidence, inference, and validity

A score is not self-interpreting. The number 17/20 does not carry its own meaning. It acquires meaning through an argument linking the task, the response, the scoring criteria, and the use made of the result. Modern validity theory therefore locates validity not in a test or score as an object, but in the interpretations and uses supported by evidence and assumptions (Kane 2013; Messick 1995). The same discipline is required of classroom grading.

The proposed interpretation of an ICPS–M score is intentionally limited:

Permitted interpretation

A score reports the degree to which a particular submitted artifact satisfied the published semantic, analytical, reasoning, and contextual criteria of a particular transfer task under the recorded conditions of production.

It does *not* by itself establish a permanent level of intelligence, motivation, mathematical identity, diligence, character, general readiness, or future potential. It does not even exhaust the student's understanding of the topic. It is time-indexed evidence from one designed encounter. Repeated artifacts may support broader claims, but those claims must be earned through patterns, not smuggled into a single score.

3.1 The evidentiary chain

The score should remain connected to its origin. In phenomenological terms, numerical grades readily become sedimented: the originating acts, distinctions, and judgments disappear while the number remains. Husserl's account of sedimentation concerns the way meanings become taken for granted after their genesis recedes from view (Husserl 1970). In a gradebook, the risk is immediate. A row of percentages can look precise while concealing incompatible origins. One score may reward completion, another correctness, another cooperation, another apparent effort. Their average is arithmetically exact and semantically indeterminate.

The proposed system keeps the chain visible.

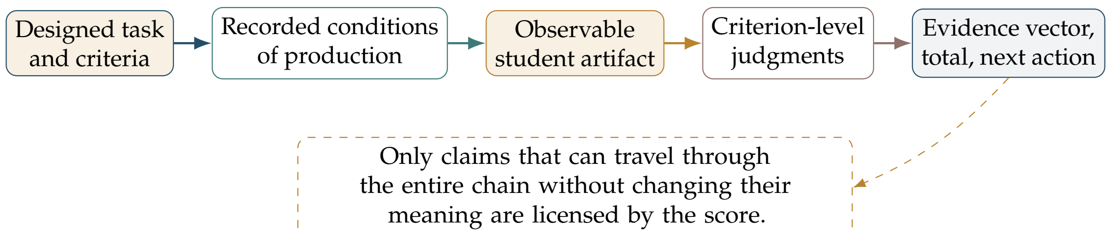

Figure 2: *The evidentiary chain prevents the score from floating free of the task, production conditions, artifact, and criterion judgments that produced it.*

Each link matters. If criteria are written after seeing the work, judgment becomes retrospective preference. If production conditions are unknown, independence cannot be inferred. If the artifact is unavailable, the score cannot be audited. If criterion judgments are collapsed into a total without trace, feedback and parent communication become impressionistic. If the report claims more than the chain supports, precision becomes pretense.

3.2 Construct-relevant and construct-irrelevant evidence

The intended construct is the student's ability, in this event, to represent and solve the mathematical problem after the specified instructional sequence. Evidence is relevant when it bears on that ability. It becomes construct-irrelevant when the score is altered by features not meant to define the performance: sociability, compliance style, handwriting aesthetics, teacher-student affinity, speed beyond the stated time condition, or willingness to speak publicly.

This is why the exclusion of participation is not a gesture of leniency. It is a validity requirement. A participation point may be easy to award, but ease is not evidence. In fact, its convenience makes it dangerous: the teacher can always find a reason after the fact, and the student can rarely determine which visible behavior would have produced a different result.

4 Why participation receives no credit

Participation is educationally important and gradationally unsuitable. The two claims are compatible. The classroom depends on attention, inquiry, responsiveness, listening, and the courage to expose uncertainty. The teacher should cultivate all of them. But a grade must not silently equate those acts with mathematical achievement.

There are four reasons.

4.1 Visible participation is multiply interpretable

The same behavior may arise from different mathematical and interpersonal conditions. Alexander (2026) treats this as a central phenomenological warning: silence may disclose confusion, resistance, concentration, fear, or advanced anticipation; rapid confidence may disclose fluency or reckless guessing. The teacher's first interpretation must therefore remain provisional. A point, however, is not provisional once it enters the gradebook.

4.2 Participation cannot be attributed cleanly

In collaborative work, one student's idea may organize the group while another records it, a third tests it, and a fourth says little but notices the contradiction that changes the plan. A single participation score cannot preserve these roles without continuous surveillance and an increasingly elaborate behavioral rubric. Such surveillance changes the activity it claims merely to observe. It also invites students to perform participation for the teacher rather than participate in the mathematical object.

4.3 Participation rewards style as if it were substance

Public verbal fluency, cultural norms of deference, language background, disability, temperament, social confidence, and prior classroom experience all shape visible participation. To award mathematical points for the visible style of engagement risks grading the form in which a person appears rather than the quality of the mathematical artifact. This is precisely the conversion from person to representation that the larger Thought Museum project resists.

4.4 Participation corrupts the interpretability of the category

Suppose two students both receive eighteen of twenty classwork points. One produced an accurate, complete artifact but spoke rarely. The other produced a partial artifact but was energetic in discussion. If participation has been blended into the score, the identical number represents different achievements. Neither student nor parent can interpret the result. The grade has purchased apparent fairness at the price of meaning.

Participation policy

> Participation is expected as a feature of the learning environment, but it is not scored. Students may be invited to speak, question, compare, listen, revise, and collaborate. None of those visible behaviors earns or loses mathematical points. Conduct is governed by classroom and school policy. Mathematical classwork is governed by the submitted artifact.

The result is not a classroom without accountability. It is a classroom in which accountability is assigned to the correct domain. A student may be corrected for disruption without pretending that the correction is a measurement of algebra. A student may be encouraged to ask questions without converting personality into points. A student may be required to keep usable notes because notes are an access tool, while the quality of those notes remains ungraded unless a separate, explicitly defined artifact makes them the object of assessment.

5 The genealogy of ICPS

The acronym ICPS is not invented for this project. It names a tradition developed by Myrna B. Shure and George Spivack: *Interpersonal Cognitive Problem Solving*. Their work examined the relation between problem-solving modes of thought and children's social adjustment and developed preventive programs for preschool and kindergarten settings (Shure and Spivack 1979, 1980, 1982; Spivack and Shure 1974). Shure later rendered the acronym in an affectionate and memorable learner-facing phrase: *I Can Problem Solve* (Shure 1993).

The present adaptation begins not by borrowing results but by attending to the design logic of Shure's dialogue. In *Thinking Parent, Thinking Child*, Shure contrasts a dialogic approach with power, suggestion, and explanation. The power approach compels an immediate outcome. The suggesting approach supplies the solution. The explaining approach supplies the adult's reasons. All three may leave the child dependent upon the adult as the owner of the problem and its resolution. Dialogue instead asks questions that return cognitive authorship to the learner (Shure 2005, pp. xv–xxiii).

The characteristic questions are deceptively simple:

- What is the problem?
- How are the people involved feeling?
- What happened before? What happened after?
- What is happening now?
- Can you think of a *different* way?
- Is that a good idea or not a good idea?
- What might happen next?

Shure emphasizes relational words such as *same*, *different*, *not*, *before*, and *after*. These words do not merely label content. They organize discriminations. They help a learner compare alternatives, trace consequences, and connect present action with a desired end. The adult is not absent; the adult structures the dialogue. But the structure is designed to help the learner figure out "how to think, not what to think" (Shure 2005, p. xxii).

5.1 The transfer of authorship

The deepest feature of the Shure dialogue is a transfer of authorship. The adult does not transfer authorship by refusing help. Nor does the adult preserve it by dictating every move. The dialogue makes the problem available in a form the learner can inhabit, then requires the learner to generate, compare, anticipate, and choose.

That architecture corresponds strikingly with cognitive apprenticeship. Modeling makes otherwise tacit expert activity visible; coaching and scaffolding support performance; articulation and reflection expose the learner's organization of the task; exploration transfers responsibility (Collins et al. 1989; Wood et al. 1976). It also corresponds with the movement described in *From Taxonomy to Tact*: the teacher organizes a field of salience, acts provisionally, observes what the intervention makes possible, and revises (Alexander 2026).

The mathematical adaptation therefore retains the dialogic transfer of authorship while changing the domain, the objects, and the evidentiary consequences.

5.2 A strict boundary of inference

What the Shure literature does not prove

The empirical and programmatic literature on Interpersonal Cognitive Problem Solving concerns social cognition, interpersonal alternatives, consequences, adjustment, and preventive intervention with children. It does not establish that the mathematical sequence developed here improves algebraic transfer, produces valid grades, or reduces grading ambiguity. Those are new claims requiring separate evidence. The present project acknowledges a genealogy, adapts a dialogic architecture, and proposes a locally testable design. It does not rename borrowed validity as innovation.

This distinction protects both traditions. Shure's work is not reduced to a mnemonic for worksheets. The mathematical design is not insulated from study by an empirical history that belongs elsewhere. Exactitude begins by refusing the rhetorical shortcut that a familiar acronym can provide.

6 ICPS–M: a mathematical adaptation

Within technical documentation, the adaptation will be designated ICPS–M: *Interpersonal Cognitive Problem Solving in Mathematics*. In student-facing language, the course may use the simpler acronym ICPS and Shure's phrase *I Can Problem Solve*. The suffix distinguishes the research object without burdening ordinary classroom speech.

The word *interpersonal* remains important. The culminating artifact is independent, but the route to it is not solitary. Mathematical understanding develops through a teacher's modeling, the student's questions, peer comparison, public representations, negotiated meanings of notation, and shared attention to error. Even when a student works alone, the mathematical practice itself is inherited and social. The final independence is therefore not isolation. It is the capacity to act responsibly within a practice that others have made available.

6.1 Adapted dialogic operators

The Shure dialogue can be translated into mathematical operators without pretending that social conflict and mathematical modeling are the same kind of problem. The correspondence is functional rather than literal.

Table 2: *From Shure's dialogic language to mathematical problem-solving prompts.*

Dialogic word or question	Cognitive function in Shure's approach	Mathematical adaptation
What is the problem?	Establishes the situation before prescribing a response	What is being asked? Which quantity, relation, object, or decision must be determined?
Same / different	Supports comparison and the generation of alternatives	What changed across the two cases? What remained invariant? Are the surface stories different while the mathematical structure is the same?

Dialogic word or question	Cognitive function in Shure's approach	Mathematical adaptation
Before / after	Connects actions with consequences across time	What was true before the transformation? What is true after it? Which property was preserved or lost?
What is happening now?	Redirects attention from escalation to the present relation	What relation is presently represented by this equation, graph, diagram, or table? What does the current line of work commit us to?
Can you think of a different way?	Generates alternatives rather than accepting the first impulse	Can the problem be represented verbally, symbolically, graphically, numerically, or geometrically? Is there another valid strategy?
Good / not good?	Evaluates an alternative through anticipated consequences	Is this representation or method fit for the stated goal and domain? What would make it invalid, inefficient, or incomplete?
What might happen next?	Develops consequential and means–ends thinking	If this assumption, operation, or model is used, what follows? Does the result remain feasible and meaningful in context?
How do you feel?	Makes the learner's lived situation available to dialogue	What seems uncertain, threatening, obvious, or confusing? What kind of help would enlarge rather than replace your action?

These prompts are not a script that must be recited. They are a repertoire. Pedagogical tact determines which question opens the field and which question merely adds noise. The teacher may need to model an entire relation before a learner can answer meaningfully; at another moment, one quiet question may be sufficient. The adaptation therefore preserves a paradox: the cycle is highly specified at the level of architecture and deliberately non-algorithmic at the level of human response.

6.2 The language of agency: "I Can Problem Solve"

The phrase *I Can Problem Solve* deserves more than motivational use. Each word carries a commitment.

The four terms of the learner-facing affirmation

I The learner is an agent, not a location at which a teacher's procedure is deposited. The first-person claim does not deny assistance; it identifies the person who must finally see, choose, execute, and answer.

Can
Capacity is real and developing. *Can* does not promise immediate success, erase prerequisite differences, or convert encouragement into evidence. It names a power that can be cultivated and demonstrated.

Problem
The problem is an object of inquiry, not a defect in the learner. Difficulty belongs to the relation among task, representation, prior knowledge, and conditions; it is not a verdict on the person.

Solve
Solving is not synonymous with producing a terminal number. It includes representing, relating, selecting, transforming, checking, interpreting, and communicating a defensible closure.

The phrase therefore sits inside the larger semantic architecture of the course. It is not the claim that every student can instantly solve every problem. It is the commitment that instruction will make the relevant structures increasingly available, that the learner will be asked to assume genuine responsibility, and that the grade will report the quality of the resulting mathematical act rather than the teacher's impression of the learner.

7 Semantic structure and mathematical problem representation

Mathematical problems do not arrive already organized for solution. They are encountered through words, diagrams, symbols, tables, graphs, conventions, and contextual assumptions. To solve is first to constitute a problem as a mathematical object: to determine what matters, how quantities and entities are related, which conditions constrain action, and what form of answer would count as closure.

Research on problem representation provides a technical basis for this claim. Experts and novices often organize problems differently. In physics, experts have been found to categorize according to underlying principles while novices rely more heavily on literal surface features (Chi, Feltovich, et al. 1981). Research on arithmetic and algebra word problems distinguishes the textbase from a situation model and a mathematical problem model; successful solution requires coordinating propositional information, situational relations, and formal operations (Kintsch and Greeno 1985; Nathan et al. 1992). Hegarty et al. (1995) further contrasts direct translation strategies, which seize numbers and key words, with problem-model strategies grounded in the described situation.

These results do not imply a single mental pipeline for all mathematics. They establish a more modest and durable point: performance depends upon how the problem is represented, and correct manipulation can coexist with a defective representation.

7.1 An operational semantic frame

For design and scoring, this paper introduces an operational notation. It is a pedagogical heuristic, not a claim that cognition literally stores problems as tuples. For a task P, define its semantic frame as

$$\mathsf{Sem}(P) = \langle \mathcal{E}, \mathcal{Q}, \mathcal{R}, \mathcal{C}, \mathcal{G}, \mathcal{V} \rangle,$$

where:

- $\mathcal{E}$ is the set of relevant entities or objects;
- $\mathcal{Q}$ is the set of quantities, variables, parameters, and units;
- $\mathcal{R}$ is the set of relations among them;
- $\mathcal{C}$ is the set of constraints, domains, conditions, and admissibility rules;
- $\mathcal{G}$ is the goal or question that determines what counts as completion;
- $\mathcal{V}$ is the available family of representational views: verbal, symbolic, numerical, graphical, tabular, diagrammatic, or physical.

The frame is useful because it separates a problem's semantic commitments from any one surface form. A student may translate a verbal relation into an equation, an equation into a graph, or a diagram into a system. A strong translation preserves relevant relations and constraints while changing the medium. A weak translation copies visible numbers while losing the situation that gives them meaning.

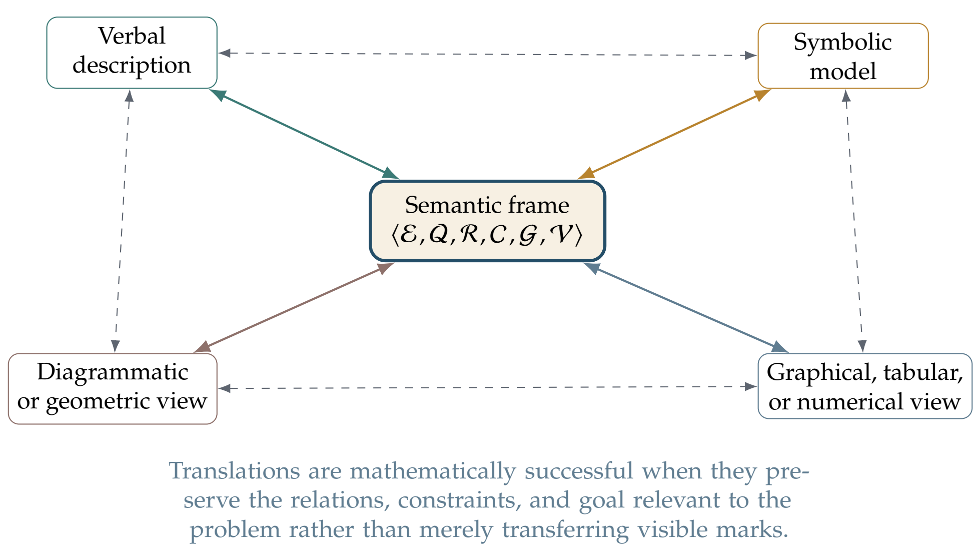

Figure 3: *The semantic frame is not another required student diagram. It is a design notation for determining what a task requires students to preserve across representations.*

7.2 Multiple representations: opportunity and burden

Multiple representations can disclose structure that one representation conceals, constrain interpretation, complement information, or support the construction of deeper understanding. They also impose cognitive tasks: learners must know how representations correspond, which information is distributed across them, and when a translation is warranted (Ainsworth 2006; Lesh et al. 1987). More representations are therefore not automatically better. An ICPS–M cycle should introduce representational variation only when it serves a defined function.

The teacher's model problem may make two or three views explicit and narrate their relation. The collaborative variation may remove one supplied view and require students to construct it. The independent transfer task may alter the surface context while preserving an underlying relational form. This sequence does not merely fade help. It changes what the learner must notice and author.

7.3 Semantic structure and the ethics of grading

A grading system that rewards only terminal correctness may miss whether the student solved the intended problem. A correct numerical result can be reached through an invalid model, a lucky cancellation, an unrecorded guess, or an inappropriate rule. Conversely, a student may construct the correct semantic model and make a local computational error. The two performances are not equivalent, and a close-grading system should not flatten them.

This is one reason the proposed rubric begins with *Semantic Representation*. The order is not decorative. It records a philosophical priority: operations acquire their mathematical meaning within a represented structure. The paper does not claim that representation always precedes calculation consciously. Experts may perceive and act fluently. But a graded artifact must make enough of the relevant relation available that the teacher can distinguish mathematical judgment from accidental arrival.

8 The ICPS–M instructional cycle

The cycle contains six phases. The first five belong to the classroom event; the sixth closes the evidentiary loop. Fascicle I supplies the single complete worked exemplar. The present section specifies the architecture without reproducing that example.

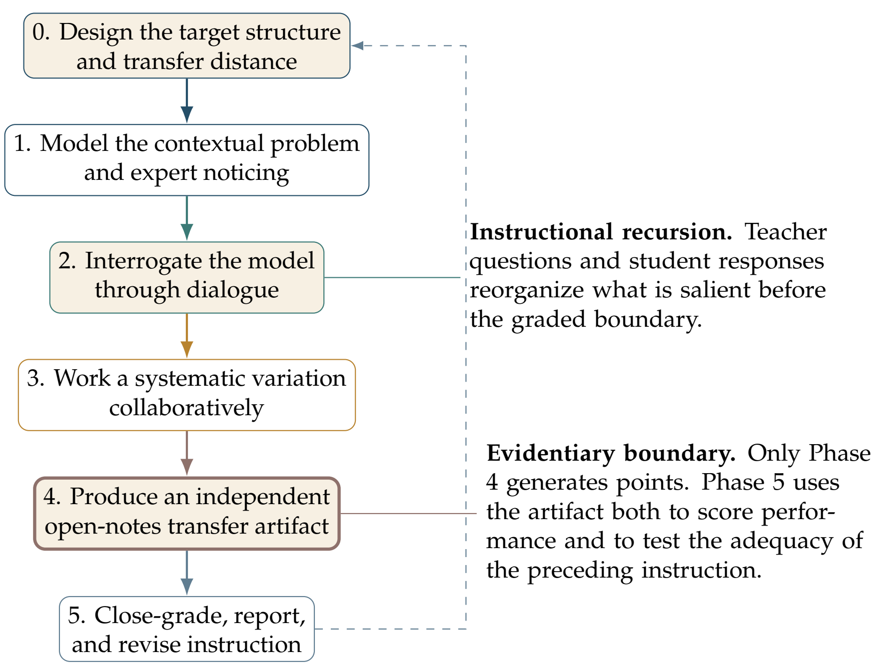

Figure 4: *The complete ICPS–M cycle. The return arrow expresses pedagogical tact: instruction is revised in light of what the artifact discloses.*

8.1 Phase 0: design before encounter

The teacher first identifies the target semantic structure, the intended cognitive processes, the expected mode of performance, and the transfer distance among the three problems. Bloom's taxonomy is useful here when it classifies the task rather than the student (Anderson and Krathwohl 2001; Bloom et al. 1956; Krathwohl 2002). The teacher specifies what students must do with what kind of knowledge: identify a relation, construct a representation, select a method, justify a decision, evaluate feasibility, or communicate a closure.

The teacher also prepares the task-specific SARC–20 rubric before student work is seen. This is non-negotiable. A rubric created afterward is not a criterion system; it is a rationalization system.

8.2 Phase 1: model the problem and expert noticing

A complete worked example can reduce unproductive search and make a procedure available to learners who do not yet possess a stable schema (Sweller and Cooper 1985). But a worked example is not self-explanatory. Students may copy operations while failing to notice why those operations were selected. Research on self-explanation shows the importance of eliciting the principles and relations that organize example use (Chi, Bassok, et al. 1989; Renkl 1997).

The model phase therefore includes more than a polished solution. The teacher externalizes expert noticing:

- what the context asks;
- which quantities vary and which remain fixed;
- which relation is decisive;
- why a representation is chosen;
- how domain restrictions arise;
- what alternative methods are possible;
- how the conclusion returns to the situation.

Students are expected to take usable notes and may ask questions. Notes are not scored. They are part of the access conditions of the later independent artifact. The teacher should design the notes students can reasonably produce, not merely display a final page too rapidly to be appropriated.

8.3 Phase 2: dialogic interrogation

The teacher then opens the model to questions. This phase is where Shure's dialogic operators become mathematically active: same/different, before/after, good/not good, and a different way. The purpose is not to hold a ritual discussion. It is to reveal whether students have noticed the structure that the model was intended to disclose.

Questions can ask students to predict the effect of changing one assumption, explain why a domain restriction matters, compare two representations, identify an attractive but invalid move, or state what the problem would become if one relation changed. Short oral responses, think-pair-share, private writing, and diagrams may all serve. None is scored. The teacher is gathering evidence for the next instructional move, not building a hidden participation ledger.

8.4 Phase 3: collaborative systematic variation

The second problem is similar enough to permit analogical comparison and different enough to expose what has actually been learned. A mere change of numbers may produce repetition without discrimination. Systematic variation instead changes selected features against a background of invariance. Variation theory emphasizes that critical aspects become discernible when learners experience variation in the relevant dimension while other aspects remain stable (Marton and Pang 2006). Analogical encoding research similarly supports comparison as a resource for transfer (Gentner et al. 2003).

The collaborative problem should therefore be designed around a declared transfer purpose. It may:

- reverse the direction of change while preserving the relational form;
- make an implicit constraint explicit or remove one that was supplied;
- introduce a discrete decision into a continuous model;
- shift the most useful representation;
- create a tie or boundary case that defeats a memorized conclusion;
- preserve the deep structure while changing the surface context.

Students may work as a class or in small groups. The teacher may coach, question, and interrupt. No group grade is assigned. The phase is a workshop in which interpretations become public and revisable.

8.5 Phase 4: independent open-notes transfer

The third problem is completed independently, normally on the same day or the next instructional day. Notes from the model and collaborative phases are permitted unless the task explicitly states otherwise. This is deliberate. The target is not unaided recall of the teacher's words. It is the student's capacity to use available records to represent and solve a related problem.

Open notes do not make the task non-independent. The notes cannot decide which relation matters, whether the new context preserves the earlier structure, which method fits the domain, or how the conclusion must be interpreted. They preserve access to prior instruction while responsibility for present judgment shifts to the student.

The production conditions are recorded:

- time available;
- permitted notes, calculator, reference sheet, and other tools;
- whether a diagram or explanation is required;
- what questions the teacher may answer after the task begins;
- whether the artifact is completed in one sitting;
- any authorized accommodation affecting access conditions.

The teacher may clarify wording but may not supply a mathematical decision that the rubric asks the student to make. When such assistance becomes necessary, it is recorded so that the score's interpretation remains honest.

8.6 Phase 5: close grading and instructional return

The artifact is scored criterion by criterion. Each lost point is attached to a public task-specific criterion or code. The resulting evidence vector is reported to the student and entered in the high-resolution ledger. The teacher then examines the distribution of errors across the class. A common semantic error may disclose a defect in the model problem or collaborative variation rather than a sudden collective failure of effort.

This return to instruction is central. *From Taxonomy to Tact* describes teaching as disciplined provisionality: act, observe what the act becomes, and revise (Alexander 2026; Manen 1991; Schön 1983). The independent artifact is not only evidence about the student. It is evidence about the instructional event.

8.7 The changing profile of support

The four instructional variables from *From Taxonomy to Tact*–structure, relational support, challenge, and autonomy–do not move in a single line. Their profile across the cycle is conceptualized in Figure 5.

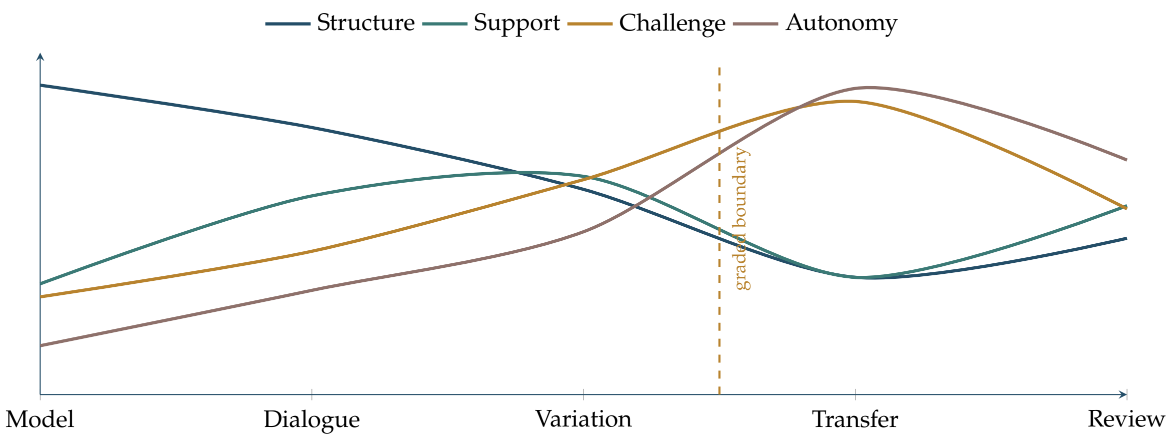

Figure 5: *Conceptual profiles across an ICPS–M cycle. Structure does not vanish; support changes form; challenge and autonomy peak in the independent transfer phase. Actual profiles remain task- and learner-sensitive.*

The figure is a design heuristic, not a psychometric scale. It prevents two symmetrical errors: making the model phase so sparse that novices cannot enter the field, and making the transfer phase so assisted that the artifact cannot support an inference of independent action.

9 The graded artifact

The final artifact is the hinge of the system. It must be rich enough to support defensible criterion judgments and bounded enough to be completed and close-graded with reasonable regularity. It is not a test in miniature, because the preceding instruction is part of its design and notes remain available. It is not practice, because the student assumes independent responsibility and the work is scored. It is not a participation product, because the earlier collaborative contribution is not converted into points.

9.1 Minimum properties

Every graded ICPS–M artifact must satisfy the following design conditions:

1. **Identifiable target.** The mathematical relation, representation, or decision being evidenced is named in the teacher design record.
2. **Specified transfer.** The relation between the final task and the model/collaborative tasks is deliberate rather than accidental.
3. **Independent authorship.** The submitted response is produced individually under recorded conditions.
4. **Auditable reasoning.** The task requires enough visible work to distinguish a defensible method from an unsupported answer.
5. **Task-specific rubric.** The allocation of all twenty points is fixed before administration.
6. **Contextual closure.** Where the task is contextual, the final response must answer the contextual question rather than terminate at an uninterpreted expression.
7. **Recoverable provenance.** The prompt, student artifact, rubric, criterion marks, and score vector can be connected after grading.

9.2 What the artifact may look like

The artifact may contain equations, prose, diagrams, graphs, tables, annotations, or a combination. No single format is required by the category. The format follows the task. A concise solution can earn full credit when it preserves all required evidence. A long solution does not earn extra credit for length. Handwriting is judged only to the extent that illegibility prevents the mathematical record from being interpreted; penmanship is not a hidden aesthetic criterion.

Alternative valid methods must receive equivalent credit. The rubric defines the mathematical obligations of the task, not the teacher's preferred path. Where two methods reveal different strengths, the scoring notes should identify how each satisfies the same criterion. The point system is meant to constrain arbitrariness, not to standardize thought into one sequence.

10 Close grading through SARC–20

Every ICPS–M artifact is scored out of twenty points using a fixed four-dimension architecture:

$$\text{SARC–20} = S_5 + A_7 + R_4 + C_4 = 20.$$

The letters name **Semantic Representation**, **Analytical Accuracy**, **Reasoning Traceability**, and **Contextual Completion**. The dimensions remain fixed across tasks; the subcriteria are task-specific.

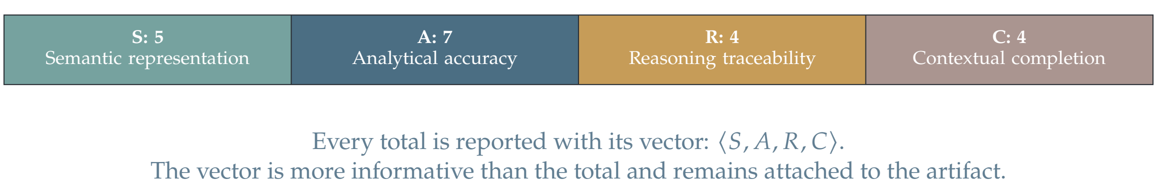

Figure 6: *The fixed SARC–20 architecture. Point values remain stable so that scores are comparable in form while task-specific descriptors preserve content validity.*

10.1 S: Semantic Representation (5 points)

Semantic Representation measures whether the artifact identifies and preserves the entities, quantities, relations, constraints, goal, and representational commitments required by the task. Depending on the problem, its five points may be assigned to variable definitions, a diagram, a relational equation, a domain, or the identification of a target quantity. The exact allocation must be visible on the task rubric.

This dimension prevents a student from receiving nearly full credit for flawless manipulation of an equation that does not represent the problem. It also protects a student who represents the problem correctly but makes a later local error: the semantic achievement remains visible rather than being erased by the final answer.

10.2 A: Analytical Accuracy (7 points)

Analytical Accuracy measures the correctness of mathematical transformations, calculations, strategy execution, and relevant use of tools. It carries the largest share because mathematics requires correctness. Yet its seven points are not reserved for matching a key. They are distributed across the actual analytical obligations of the task.

Carry-through credit is permitted. If a student makes one semantic or arithmetic error and then reasons correctly from the resulting expression, later analytical points are awarded when the corresponding work is mathematically sound. The original error is scored where it occurred. This avoids punishing one error repeatedly while preserving the truth that the final result is affected.

10.3 R: Reasoning Traceability (4 points)

Reasoning Traceability measures whether a competent reader can determine why the operations and decisions belong together. It is not a prose quota. A well-labeled diagram and a short equation chain may be fully traceable. A page of unexplained algebra may not be.

The dimension answers a central problem raised in *From Taxonomy to Tact*: successful compliance can masquerade as understanding (Alexander 2026). The artifact must disclose enough of the student's path to distinguish selection from recitation. It need not disclose private thought or require a confessional explanation. It requires a public mathematical record.

10.4 C: Contextual Completion (4 points)

Contextual Completion measures whether the response returns to the actual question with appropriate interpretation, units, domain checks, rejected or extraneous solutions, feasibility, and a clearly stated conclusion. A symbolic result is not complete when the problem asks for a price, time, dimension, rate, decision, or comparison.

This dimension also protects against ritualized "answer statements". The criterion is not the presence of a sentence stem. It is whether the mathematical work has been reconnected to the context that gave it meaning.

10.5 Generic descriptor architecture

The fixed dimensions require task-specific descriptors. The generic architecture below establishes the quality continuum without pretending that one descriptor can score every mathematical domain.

Table 3: *Generic SARC quality descriptors. The task rubric converts these obligations into exact point-bearing criteria.*

Dimension	Complete evidence	Substantial evidence	Partial evidence	Insufficient evidence
Semantic Representation	Relevant quantities, relations, constraints, and goal are represented correctly and coherently	Core structure is correct with a local omission or imprecision	Some relevant elements are present, but the controlling relation is incomplete or unstable	Work does not represent the assigned problem or leaves the target indeterminate
Analytical Accuracy	Required operations and decisions are mathematically correct	Method is appropriate and mostly correct; local errors do not destroy the main analysis	A plausible method begins but contains major execution gaps or contradictions	No assessable method, or method is fundamentally inapplicable
Reasoning Traceability	Selection and progression are visible enough to audit without unnecessary verbosity	Main reasoning is visible, with one important link left implicit	Steps are present but their relation or purpose is difficult to determine	Answer is unsupported or work is too fragmentary to reconstruct

Dimension	Complete evidence	Substantial evidence	Partial evidence	Insufficient evidence
Contextual Completion	Result is interpreted, checked, feasible, and expressed with required units or restrictions	Conclusion answers the main question but omits one contextual obligation	A numerical or symbolic result is present with weak or incomplete interpretation	No answer to the contextual question or conclusion contradicts the model

The descriptors are not themselves the scoring key. An analytic rubric improves consistency when it is content-specific, supported by exemplars, and used with calibration; a rubric does not automatically make an inference valid (Jönsson and Svingby 2007). Each ICPS–M task therefore requires a concrete allocation of the five, seven, four, and four points.

11 Scoring rules that eliminate hidden judgment

Near-zero ambiguity requires procedural rules as well as dimensions. The following rules govern every ICPS–M artifact unless a task explicitly announces a stricter condition.

11.1 Criteria precede work

All point-bearing criteria are written before administration. Criteria may be clarified if wording proves ambiguous, but they may not be expanded after seeing an undesirable response. If the task failed to specify an obligation, the failure belongs to design, not to the student.

11.2 Every lost point receives a reason

A score below twenty must be traceable to a criterion code or a brief annotation. "Incomplete", "careless", and "needs work" are not sufficient unless the rubric defines what evidence was absent or incorrect. The annotation identifies the mathematical location of the gap.

11.3 No double penalty for one originating error

A single error is scored at its origin. Correct subsequent reasoning from an incorrect intermediate result receives carry-through credit where the later criterion can still be evaluated. A later criterion receives no credit only when the original error makes the required evidence genuinely absent, not merely because the final number is wrong.

11.4 Alternative methods receive equivalent credit

No criterion may require the teacher's method unless the method itself is the stated learning target. A student who supplies a valid alternative must be scored according to the same semantic, analytical, reasoning, and contextual obligations. When the alternative produces evidence in a different order or form, the grader maps it to the rubric rather than forcing the artifact into the key.

11.5 Correctness does not erase the need for evidence

A correct unsupported answer earns the points warranted by what is visible, not the points that might have been earned by an imagined method. This is not punishment for concision. It is a limit on inference. The grader cannot award evidence that the artifact does not contain.

11.6 Verbosity, neatness, and stylistic conformity are not scored

The rubric may require legibility, labels, or organization when they are necessary to interpret the mathematics. It may not reward decorative length, preferred handwriting, color, or exact imitation of the teacher's layout. The student is responsible for an auditable artifact, not an aesthetic replica.

11.7 Access conditions are recorded, not moralized

Permitted notes, calculators, reference sheets, extended time, and other authorized accommodations are conditions of production. They are not deductions. If an accommodation alters a criterion legitimately under a school plan, the altered criterion is documented in advance. If the mathematical standard remains the same, the score is interpreted under the recorded access condition.

11.8 Administrative status is not academic evidence

An absence, an unprocessed make-up, or a missing scan is not yet a mathematical zero. The ledger distinguishes **Pending**, **Scored**, and **Closed** statuses. A zero enters the academic record only when the school's completion policy has been applied or when the submitted artifact contains no evidence for the criteria. This separation prevents logistics from disguising themselves as mathematics.

The audit question

For any point awarded or withheld, the school should be able to ask: *Which public criterion does this point represent, where is the corresponding evidence in the artifact, and under what conditions was the artifact produced?* If the answer cannot be given, the point is not ready to enter the gradebook.

12 High-resolution tracking

A single percentage hides the profile that produced it. Two students may both earn eighty percent while needing entirely different next actions. High-resolution tracking preserves the criterion vector.

For student i on artifact j, let

$$\mathbf{v}_{ij} = \langle S_{ij}, A_{ij}, R_{ij}, C_{ij} \rangle, \qquad T_{ij} = S_{ij} + A_{ij} + R_{ij} + C_{ij},$$

with bounds $0 \le S_{ij} \le 5$, $0 \le A_{ij} \le 7$, and $0 \le R_{ij}, C_{ij} \le 4$. If all completed ICPS–M artifacts are equally weighted, the live Classwork percentage after n scored artifacts is

$$G_i(n) = 100 \left(\frac{\sum_{j=1}^{n} T_{ij}}{20n} \right).$$

Pending artifacts are excluded from n until the applicable completion process has concluded. The formula prevents the category from being altered by unannounced weighting. If an exceptional task is to carry a different weight, the exception must be disclosed before administration and visible in the ledger.

12.1 The evidence ledger

Each artifact record contains the following fields:

Table 4: *Minimum high-resolution ledger fields.*

Field	Meaning
ICPS identifier	Stable code connecting prompt, instructional cycle, rubric, artifact, and gradebook entry
Date and unit	Temporal and curricular location of the evidence
Target semantic structure	Relations, constraints, representations, and goal the task was designed to disclose
Transfer designation	Declared relation between model, collaborative variation, and independent artifact
Production conditions	Time, notes, tools, accommodations, and recorded assistance
SARC vector	Criterion totals $\langle S, A, R, C \rangle$
Criterion codes	Exact reasons for lost points or flagged strengths
Total	Sum out of twenty
Status	Assigned, Pending, Scored, or Closed
Next action	One concise mathematical action warranted by the evidence
Communication record	Date and channel by which the result or pattern was made available to student and, when appropriate, parent/guardian

The ledger is not meant to create a second gradebook. It is the semantic layer that makes the gradebook interpretable. A conventional entry might show "ICPS 4: 16/20". The high-resolution record shows $\langle 4, 6, 3, 3 \rangle$, the criterion codes that explain the four lost points, and the next action. The total answers *how much*; the vector answers *where*; the annotation answers *why*; the next action answers *what now*.

12.2 Student and parent communication

Students should receive the SARC vector whenever the score is returned. Parent communication should aggregate rather than merely repeat. A high-resolution update can report:

- the current Classwork percentage;
- the number of scored and pending artifacts;
- the current average within each SARC dimension;
- recurring criterion codes;
- the most recent next action;
- any recovery obligation generated by absence or noncompletion.

The Communication Journal may eventually serve as the channel for periodic summaries, while the marked artifact remains the primary evidence. The communication system should not flood families with data that cannot be interpreted. Its purpose is resolution, not volume.

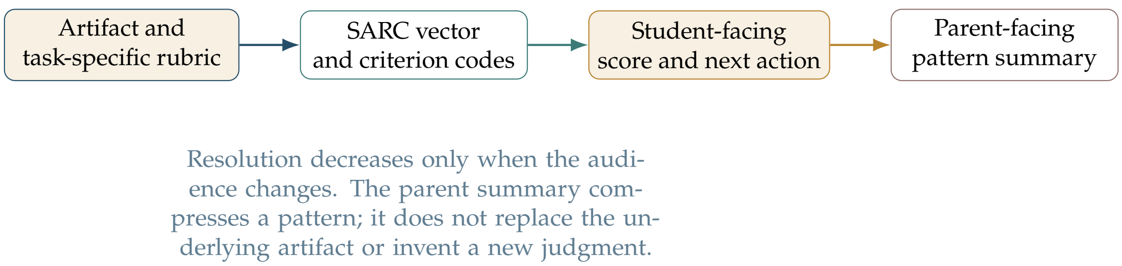

Figure 7: *High-resolution reporting preserves a trace from the artifact to the family-facing summary.*

13 Task construction and transfer calibration

The integrity of the grade depends at least as much on task design as on scoring. A precise rubric cannot rescue a task that fails to elicit the intended evidence. The ICPS–M designer therefore constructs the three problems as a coordinated set.

13.1 Model problem

The model problem should contain the target semantic structure in a context rich enough to make interpretation necessary but bounded enough to support complete exposition. Its numbers, diagrams, and wording should not create irrelevant difficulty. The teacher's solution should reveal not only operations but the basis of selection.

13.2 Collaborative variation

The collaborative problem should vary a critical aspect. Its purpose is diagnostic and generative. It should make the first solution insufficient as a script while allowing the underlying relation to be recognized. The class compares, tests, and revises. The teacher watches for whether students are translating surface features or preserving structure.

13.3 Independent transfer

The independent task should require the target structure without being a disguised copy. Transfer can be thought of along several dimensions:

- **surface distance**: change in story, vocabulary, or visual presentation;
- **representational distance**: change in which view is supplied or must be constructed;
- **relational distance**: change in the organization of quantities and constraints;
- **procedural distance**: change in the method most efficient for analysis;
- **decision distance**: change in what counts as a complete or admissible answer.

The task should not maximize all five. Transfer that is too small measures imitation; transfer that is too large may measure unintroduced knowledge. Preparation for future learning is a useful counterpoint to a simple near/far dichotomy: instruction can prepare learners to use resources in new situations rather than merely reproduce a trained response (Bransford and Schwartz 1999). The teacher's design record should therefore state what kind of transfer is intended and why it is reasonable after the preceding phases.

13.4 Worked examples, fading, and principled prompts

Worked-example research supports transitions from example study to problem solving, particularly when fading is combined with prompts that direct attention to underlying principles (Atkinson et al. 2003). ICPS–M resembles that transition but is not reducible to step fading. The collaborative problem may preserve all steps while altering the context; the independent problem may allow notes while requiring a new representation. What fades is not merely ink on a page. What changes is the locus of responsibility.

This is a crucial correction. Support can remain present while autonomy increases. Notes, a familiar rubric, and a stable problem-solving language may remain available even as the student becomes responsible for choosing and executing the solution. Autonomy is supported responsibility, not abandonment (Alexander 2026).

14 Calibration and quality assurance

A grading architecture reaches professionalism only when it can inspect itself. The following quality-assurance procedures are part of the design, not optional embellishments.

14.1 Pre-administration review

Before use, the teacher checks:

1. Does the final task actually require the target semantic structure?
2. Can every point be located in a task-specific criterion?
3. Are alternative valid methods foreseeable and scoreable?
4. Is the transfer distance deliberate and reasonable?
5. Are access conditions sufficient and explicit?
6. Can the artifact be close-graded within the promised turnaround?
7. Does any criterion unintentionally reward participation, style, or teacher preference?

14.2 Exemplar calibration

The teacher should maintain a small set of anonymized or constructed artifacts representing boundary cases for each task: complete, substantial, partial, and insufficient evidence. Exemplars do not replace criteria, but they reveal where criteria are vague. Analytic rubrics are more reliable when supported by task-specific exemplars and rater calibration (Jönsson and Svingby 2007).

For a single-teacher course, calibration can include rescoring a sample after a delay and checking whether criterion decisions remain stable. For common tasks across sections, a second reader can independently score a sample. Disagreement is not embarrassing data. It is precisely the data needed to improve the rubric.

14.3 Post-administration audit

After grading, the teacher inspects:

- which criteria produced the most lost points;
- whether one wording feature generated unintended errors;
- whether alternative methods were scored consistently;
- whether a criterion depended on information not stated in the prompt;
- whether the scoring time was operationally sustainable;
- whether students and parents could interpret the returned record.

The task version is then archived with revision notes. The next use is not treated as identical merely because the title is unchanged. This is design-based professional practice: each cycle produces evidence about both learner performance and the instrument that elicited it.

15 Pedagogical tact and scoring accountability

The demand for near-zero ambiguity could be misunderstood as a demand to automate teaching. It is not. No rubric can determine which question should be asked during the model phase, whether a student's silence requires protection or challenge, how long to wait, when to expose a contradiction, or what a particular error presently means. Those decisions belong to pedagogical tact (Alexander 2026; Manen 1991).

The architecture instead separates two forms of responsibility:

Instructional responsiveness
Interpret the learner–task relation
Adjust structure, support, challenge, autonomy
Act provisionally and revise

Scoring accountability
Apply criteria fixed in advance
Point to evidence in the artifact
Report criterion decisions and conditions

Common ethical center: the learner is never reduced
to the teacher's first representation of the learner.

Figure 8: *Tact governs responsive instruction; auditability governs scoring. The two are mutually corrective rather than mutually exclusive.*

Instruction may be singular and responsive. Scoring must be public and repeatable enough to defend. The teacher may alter the route to the artifact without secretly altering what the artifact must show. If an instructional change legitimately changes the evidentiary task, the rubric and conditions are changed openly before work is judged.

This distinction also protects the teacher. A transparent system reduces the pressure to justify a score through retrospective narrative. The teacher need not say, "I know this student and felt the work deserved more". Nor must the teacher defend an unexplained deduction by invoking rigor. The evidence chain carries the judgment.

16 Ethical limits and failure modes

No architecture eliminates misuse. Exact categories can become instruments of domination when their exactness is mistaken for completeness. Lewis's distinction between initiation and conditioning is relevant here. The system should initiate learners into a practice in which teacher and student are answerable to mathematical relations and public criteria. It should not condition students to perform preferred behaviors whose purpose remains external to them (Lewis 2001).

Several failure modes require active resistance.

16.1 Rubric literalism

A student may satisfy the mathematical purpose in a form the rubric writer did not anticipate. The grader must interpret the criterion, not worship its first wording. The remedy is not discretionary bonus points but a documented recognition that the criterion has been satisfied by an equivalent form.

16.2 Pseudo-precision

Twenty points can create the appearance of measurement finer than the evidence supports. The system answers this by attaching each point to an observable criterion and by preserving the vector. It does not claim interval-scale precision about a person. The total is a course accounting device built from defensible ordinal and categorical judgments.

16.3 Over-surveillance

High-resolution tracking may tempt the teacher to record every hesitation and behavior. That would recreate participation grading under a different name. The ledger records the artifact, production conditions, criterion decisions, and next action. It does not become a dossier of personality.

16.4 Feedback overload

Close grading can produce so many annotations that the student cannot identify the controlling issue. The system therefore requires one concise next action in addition to criterion codes. Resolution means identifying the structure of the gap, not multiplying red ink.

16.5 Misuse of absence codes

Pending status must not become indefinite invisibility. The course needs a clear recovery process aligned with school policy. The present paper establishes the semantic distinction between pending and zero; the operational calendar must be fixed in the course implementation.

16.6 Treating the vector as a new identity

A recurring low Semantic Representation average is evidence of a performance pattern under a family of tasks. It is not the statement “this student is a poor modeler”. The pattern guides instruction and practice; it does not become an ontology of the learner. Dreyfus designations and SARC profiles alike remain task- and time-specific hypotheses (Alexander 2026; Benner 1984; H. L. Dreyfus and S. E. Dreyfus 1986; S. E. Dreyfus and H. L. Dreyfus 1980).

17 Implementation protocol

The following protocol is recommended for the first year of deployment.

17.1 Before the term

1. Establish the fixed SARC–20 architecture in the course documentation.
2. Prepare the evidence-ledger template and criterion-code convention.
3. Construct at least one complete ICPS–M cycle for the opening unit.
4. Verify that the gradebook can display each twenty-point total without hidden weighting.
5. Determine the parent-facing reporting cadence and the role of the Communication Journal.
6. Establish the absence/recovery timeline in conformity with school procedures.

17.2 For each cycle

1. Complete the Phase 0 design record.
2. Publish or display the task-specific SARC rubric before the independent phase.
3. Conduct the model, dialogue, and collaborative variation without correctness points.
4. Administer the independent artifact under recorded conditions.
5. Close-grade using criterion codes and the no-double-penalty rule.
6. Return the score vector and one next action.
7. Enter total, vector, status, and communication record.
8. Audit class-level patterns and revise the cycle documentation.

17.3 At reporting intervals

The teacher reviews both the total category standing and the four-dimensional profile. Students should know not only the percentage but the evidence from which it is composed. Parents should receive a concise pattern summary when reporting is scheduled or when a defined concern threshold is reached. The exact threshold belongs in the implementation policy, not in private intuition.

A standard of administrative defensibility

At any reporting date, the Classwork grade should be reproducible from the scored artifact set; each artifact should be traceable to a published rubric; each deduction should be traceable to evidence; and each pending item should have an explicit status. The grade should not depend upon a teacher remembering how engaged the student seemed three weeks earlier.

18 Local evaluation and research agenda

The design should be studied during implementation. The following questions are especially important.

18.1 Scoring stability

When the same artifact is rescored after a delay, do criterion decisions remain stable? Where instability appears, is the problem the rubric, the artifact, or the construct? A small reliability study can be conducted without pretending that classroom scoring is a standardized test.

18.2 Transfer quality

Do students increasingly preserve deep relations across changed surface contexts? Are they relying less on keywords and more on problem models? Think-aloud interviews, error analysis, and comparison of representations can provide evidence, but such inquiry should be limited and non-intrusive (Chi, Feltovich, et al. 1981; Hegarty et al. 1995).

18.3 Instructional diagnosis

Do class-level SARC patterns help distinguish representational difficulties from execution difficulties? Does the vector produce better next actions than a single total? This is a practical validity question: whether the reporting system improves the decisions made from scores (Kane 2013).

18.4 Student interpretability

Can students explain why they received a score and what would have changed it? Can they use the rubric while producing work rather than only after grading? Sadler (1989) argues that improvement requires learners to understand quality, compare present work with a standard, and possess strategies for closing the gap. The SARC system should be judged partly by whether it builds that evaluative expertise.

18.5 Parent interpretability

Do parent summaries clarify standing or create a new layer of technical language? The report should be tested for whether families can identify the mathematical strength, the current gap, and the next action without needing to decode a private system.

18.6 Equity and unintended consequences

Does the exclusion of participation reduce construct-irrelevant variance? Do the independent artifacts create new barriers through language, time, or unfamiliar contexts? Are accommodations functioning as intended? Exact scoring is not self-justifying; consequences remain part of the validity argument (Messick 1995).

18.7 Boundary of current warrant

Until these questions are studied, the strongest justified claim is that the design creates a more explicit evidentiary chain and a more auditable grading procedure. Claims about achievement gains, long-term transfer, motivation, or social-emotional outcomes must await evidence. The paper's confidence lies in the exactness of the design, not in the inflation of its results.

19 Compression-ready propositions

The eventual syllabus policy should be short. It should not attempt to reproduce this paper. It should, however, be a faithful compression of the following propositions.

1. Classwork is twenty percent of the course grade and consists only of scored ICPS–M independent artifacts.
2. Participation, behavior, note-taking, group contribution, apparent effort, and preparedness do not earn mathematical points.
3. Each ICPS–M cycle includes a modeled contextual problem, detailed teacher exposition, dialogic questioning, a collaborative variation, and a final independent transfer problem.
4. The first phases are instruction and are not graded for correctness.
5. The final artifact is ordinarily open notes, supervised, and independently authored.
6. Every artifact is worth twenty points under the fixed SARC architecture: Semantic Representation (5), Analytical Accuracy (7), Reasoning Traceability (4), and Contextual Completion (4).
7. Task-specific point criteria are established before work is judged.
8. Valid alternative methods receive equivalent credit.
9. One originating error is not penalized repeatedly; mathematically correct carry-through work receives the points it evidences.
10. Every lost point is connected to a criterion; no hidden participation, neatness, effort, or impression points enter the total.

11. Students receive the total, SARC vector, criterion annotations, and one next action.
12. Pending absences and administrative noncompletion are coded separately from academic zeroes until the applicable process concludes.
13. Parent reporting preserves a trace to the underlying artifacts and summarizes criterion patterns rather than teacher impressions.
14. Scores describe performances under specified conditions, not permanent traits or identities.
15. The instructor audits recurring error patterns and revises instruction and task design accordingly.

These propositions are sufficient to draft a dense student-facing abstract later. They are not themselves the final wording. The final wording must also fit the school's platform, parent communication practice, and absence procedures.

20 Conclusion: what a classwork point is for

The ordinary classroom is full of things worth seeing that are not worth scoring. A question may open an entire field. A silence may protect thought. A wrong proposal may reveal the decisive distinction. A group may discover a relation that no individual had yet articulated. These events belong to teaching. Their value would be diminished, not increased, if each were translated into points.

The Classwork category should therefore score less and mean more. It should score one defined artifact produced after a serious instructional sequence. That artifact should be rich enough to show how the problem was represented, how the mathematics was executed, how the reasoning can be traced, and how the result returns to the context. Its score should remain attached to criteria, conditions, and evidence. Its limitations should remain visible. Its pattern should inform the next act of teaching.

The proposed ICPS–M architecture joins Shure's transfer of problem-solving authorship to a mathematical account of semantic structure, worked-example transition, systematic variation, independent transfer, and auditable judgment. The learner-facing phrase *I Can Problem Solve* then becomes neither sentiment nor slogan. It names the purpose of the sequence: the emergence of a learner who can make a problem intelligible, act within its relations, and assume responsibility for a defensible solution.

Near-zero ambiguity does not mean that a machine could replace the teacher. It means that the teacher's indispensable judgment is exercised where it belongs. During instruction, judgment remains tactful, responsive, and provisional. During scoring, judgment becomes public, criterion-bound, and answerable to evidence. The system does not eliminate the teacher. It eliminates the unexplained point.

References

Ainsworth, S. (2006). "DeFT: A Conceptual Framework for Considering Learning with Multiple Representations". *Learning and Instruction* 16.3, pp. 183–198. doi: 10.1016/j.learninstruc.2006.03.001.

Alexander, P. (Aug. 2026). *From Taxonomy to Tact: A Phenomenological Synthesis of Bloom's Taxonomy, the Dreyfus Model, and Situational Leadership in Education*. Instructional Theory Series, Version 3. Thought Museum.

Anderson, L. W. and D. R. Krathwohl, eds. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.

Atkinson, R. K., A. Renkl, and M. M. Merrill (2003). "Transitioning from Studying Examples to Solving Problems: Effects of Self-Explanation Prompts and Fading Worked-Out Steps". *Journal of Educational Psychology* 95.4, pp. 774–783. doi: 10.1037/0022-0663.95.4.774.

Benner, P. (1984). *From Novice to Expert: Excellence and Power in Clinical Nursing Practice*. Menlo Park, CA: Addison-Wesley.

Bloom, B. S., M. D. Engelhart, E. J. Furst, W. H. Hill, and D. R. Krathwohl (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain*. New York: David McKay.

Bransford, J. D. and D. L. Schwartz (1999). "Rethinking Transfer: A Simple Proposal with Multiple Implications". *Review of Research in Education* 24, pp. 61–100. doi: 10.3102/0091732X024001061.

Chi, M. T. H., M. Bassok, M. W. Lewis, P. Reimann, and R. Glaser (1989). "Self-Explanations: How Students Study and Use Examples in Learning to Solve Problems". *Cognitive Science* 13.2, pp. 145–182. doi: 10.1207/s15516709cog1302_1.

Chi, M. T. H., P. J. Feltovich, and R. Glaser (1981). "Categorization and Representation of Physics Problems by Experts and Novices". *Cognitive Science* 5.2, pp. 121–152. doi: 10.1207/s15516709cog0502_2.

Collins, A., J. S. Brown, and S. E. Newman (1989). "Cognitive Apprenticeship: Teaching the Crafts of Reading, Writing, and Mathematics". *Knowing, Learning, and Instruction: Essays in Honor of Robert Glaser*. Ed. by L. B. Resnick. Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 453–494.

Dreyfus, H. L. and S. E. Dreyfus (1986). *Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*. New York: Free Press.

Dreyfus, S. E. and H. L. Dreyfus (1980). *A Five-Stage Model of the Mental Activities Involved in Directed Skill Acquisition*. ORC 80-2. Operations Research Center, University of California, Berkeley.

Gentner, D., J. Loewenstein, and L. Thompson (2003). "Learning and Transfer: A General Role for Analogical Encoding". *Journal of Educational Psychology* 95.2, pp. 393–408. doi: 10.1037/0022-0663.95.2.393.

Hegarty, M., R. E. Mayer, and C. A. Monk (1995). "Comprehension of Arithmetic Word Problems: A Comparison of Successful and Unsuccessful Problem Solvers". *Journal of Educational Psychology* 87.1, pp. 18–32. doi: 10.1037/0022-0663.87.1.18.

Husserl, E. (1970). *The Crisis of European Sciences and Transcendental Phenomenology: An Introduction to Phenomenological Philosophy*. Trans. by D. Carr. Evanston, IL: Northwestern University Press.

Jönsson, A. and G. Svingby (2007). "The Use of Scoring Rubrics: Reliability, Validity and Educational Consequences". *Educational Research Review* 2.2, pp. 130–144. doi: 10.1016/j.edurev.2007.05.002.

Kane, M. T. (2013). "Validating the Interpretations and Uses of Test Scores". *Journal of Educational Measurement* 50.1, pp. 1–73. doi: 10.1111/jedm.12000.

Kintsch, W. and J. G. Greeno (1985). "Understanding and Solving Word Arithmetic Problems". *Psychological Review* 92.1, pp. 109–129. doi: 10.1037/0033-295X.92.1.109.

Krathwohl, D. R. (2002). "A Revision of Bloom's Taxonomy: An Overview". *Theory Into Practice* 41.4, pp. 212–218. doi: 10.1207/s15430421tip4104_2.

Lesh, R., T. Post, and M. Behr (1987). "Representations and Translations among Representations in Mathematics Learning and Problem Solving". *Problems of Representation in the Teaching and Learning of Mathematics*. Ed. by C. Janvier. Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 33–40.

Lewis, C. S. (2001). *The Abolition of Man, or, Reflections on Education with Special Reference to the Teaching of English in the Upper Forms of Schools*. New York: HarperOne.

Manen, M. van (1991). *The Tact of Teaching: The Meaning of Pedagogical Thoughtfulness*. Albany: State University of New York Press.

Marton, F. and M. F. Pang (2006). "On Some Necessary Conditions of Learning". *Journal of the Learning Sciences* 15.2, pp. 193–220. doi: 10.1207/s15327809jls1502_2.

Messick, S. (1995). "Validity of Psychological Assessment: Validation of Inferences from Persons' Responses and Performances as Scientific Inquiry into Score Meaning". *American Psychologist* 50.9, pp. 741–749. doi: 10.1037/0003-066X.50.9.741.

Nathan, M. J., W. Kintsch, and E. Young (1992). "A Theory of Algebra-Word-Problem Comprehension and Its Implications for the Design of Learning Environments". *Cognition and Instruction* 9.4, pp. 329–389. doi: 10.1207/s1532690xci0904_2.

Renkl, A. (1997). "Learning from Worked-Out Examples: A Study on Individual Differences". *Cognitive Science* 21.1, pp. 1–29. doi: 10.1207/s15516709cog2101_1.

Sadler, D. R. (1989). "Formative Assessment and the Design of Instructional Systems". *Instructional Science* 18.2, pp. 119–144. doi: 10.1007/BF00117714.

Schön, D. A. (1983). *The Reflective Practitioner: How Professionals Think in Action*. New York: Basic Books.

Shure, M. B. (1993). "I Can Problem Solve (ICPS): Interpersonal Cognitive Problem Solving for Young Children". *Early Child Development and Care* 96, pp. 49–64. doi: 10.1080/0300443930960106.

– (2005). *Thinking Parent, Thinking Child: How to Turn Your Most Challenging Everyday Problems into Solutions*. New York: McGraw-Hill.

Shure, M. B. and G. Spivack (1979). "Interpersonal Cognitive Problem Solving and Primary Prevention: Programming for Preschool and Kindergarten Children". *Journal of Clinical Child Psychology* 8.2, pp. 89–94. doi: 10.1080/15374417909532894.

– (1980). "Interpersonal Problem Solving as a Mediator of Behavioral Adjustment in Preschool and Kindergarten Children". *Journal of Applied Developmental Psychology* 1.1, pp. 29–44. doi: 10.1016/0193-3973(80)90060-X.

– (1982). "Interpersonal Problem-Solving in Young Children: A Cognitive Approach to Prevention". *American Journal of Community Psychology* 10.3, pp. 341–356. doi: 10.1007/BF00896500.

Spivack, G. and M. B. Shure (1974). *Social Adjustment of Young Children: A Cognitive Approach to Solving Real-Life Problems*. San Francisco: Jossey-Bass.

Sweller, J. and G. A. Cooper (1985). "The Use of Worked Examples as a Substitute for Problem Solving in Learning Algebra". *Cognition and Instruction* 2.1, pp. 59–89. doi: 10.1207/s1532690xci0201_3.

Wood, D., J. S. Bruner, and G. Ross (1976). "The Role of Tutoring in Problem Solving". *Journal of Child Psychology and Psychiatry* 17.2, pp. 89–100. doi: 10.1111/j.1469-7610.1976.tb00381.x.