Writing Classes for Mathematical Modeling Competitions

W2: The Summary – Claim, Evidence, and Decision

Kenneth Sok Kin Cheng

Revised 20 August 2026

Learning goals

By the end of this two-hour writing class, students should be able to:

- distinguish a competition **Summary Sheet** from an introduction, abstract, methods list, or team diary;
- build an **evidence inventory** before drafting prose;
- map every major task in the problem statement to a model, result, and decision;
- organize a summary through four functional blocks: task, strategy, findings, and trust/decision;
- select a small set of decision-relevant numbers and report them with units, baselines, and conditions;
- distinguish verification, validation, sensitivity, and uncertainty in one-sentence summary claims;
- write conditional recommendations without weakening the paper;
- compress a draft while preserving task coverage and evidence;
- conduct a rapid peer audit using answer visibility, evidence, consistency, and precision.

1 What a competition summary must accomplish

The Summary is not a miniature version of every section. It is a decision brief: a reader should be able to identify the task, the modeling route, the main findings, and the confidence or limitation attached to the recommendation without searching the full report.

Official guidance supports this emphasis. The current HiMCM instructions require the solution to begin with the Summary Sheet and ask teams to present major ideas and results clearly. The current IM^{2}C USA instructions state that judges place considerable weight on the summary and that the most important conclusions should be prominent. Neither source supports invented claims such as "judges spend exactly seven minutes" or "the summary determines 60% of the score." Use official expectations, not folklore.

Launch activity: the 20-second answer test

Read a summary page for twenty seconds, then close it. Without looking back, answer:

1. What decision or quantity was the team asked to produce?
2. What was the team's main recommendation or numerical answer?
3. What model or method generated it?
4. What evidence makes the recommendation credible?

If a reader cannot recover at least three answers, the summary is not yet functioning as a summary.

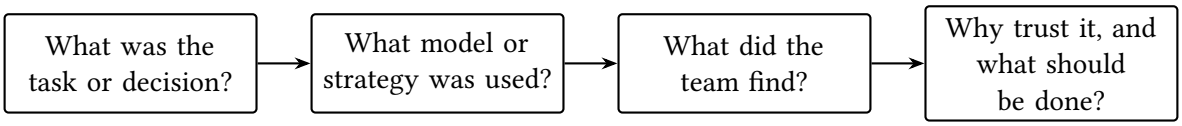

Figure 1: The summary should answer four reader questions. The wording and order may vary, but the functions should remain visible.

Model criticism: Summary, introduction, and abstract are not identical

- An **introduction** develops context, problem interpretation, and report organization. It may postpone numerical answers.
- A research **abstract** often follows disciplinary conventions and may emphasize objective, method, result, and conclusion in compressed form.
- A competition **Summary Sheet** is closer to an executive decision brief: it should state the answer explicitly, identify the modeling architecture, and indicate why the answer is defensible.

Copying the introduction into the Summary usually produces too much background and too few results.

2 Do not begin with prose: build the evidence inventory

A weak summary is often not a sentence-level problem. It is an evidence-selection problem. Teams begin writing before deciding which tasks, results, comparisons, and limitations must survive the one-page constraint.

The five-column evidence inventory

Before drafting, complete one row for every important task:

Task	Headline answer	Method	Evidence/check	Condition/caveat
What must be answered?	Number, ranking, route, threshold, or recommendation	Model or algorithm that produced it	Baseline, validation, uncertainty, or sensitivity	When might the answer change?

A row with no headline answer signals unfinished analysis. A row with no evidence signals an unsupported claim. A row with no condition may signal overconfidence.

Worked example 1: evidence inventory for an illustrative school-bus case

A school must redesign routes for 226 students using buses of capacity 60. The target is to reduce ride time while keeping the longest scheduled ride below 45 minutes.

Task	Answer	Method	Evidence/check	Condition
Choose fleet size	Four buses	Capacity lower bound plus routing model	$226/60 = 3.77$, so at least four are required	Assumes all registered riders travel that day
Design routes	Four routes, 87.4 km total	Cluster-first vehicle-routing heuristic with local search	Existing routes total 103.6 km	Road closures excluded
Reduce ride time	Mean 28.2 min; maximum 44.1 min	Time-dependent shortest paths	Existing mean 42.9 min; all planned routes below 45 min	Uses morning travel-time estimates
Test demand changes	Four buses sufficient in 18 of 20 tested demand scenarios	Reassignment under ± 10 students per district	Maximum ride remains below 45 min in 17 scenarios	A fifth bus is needed in two high-demand scenarios

The Summary does not need every number in this table. It needs the numbers that answer the task, establish improvement, and reveal the decision boundary.

Checkpoint: Checkpoint: what is missing?

A team writes: "Our optimized route length is 87.4 km." Identify at least three missing pieces from the evidence inventory. Possible questions include: compared with what, for how many buses, under which capacity and time constraints, and how stable is the result?

Model criticism: Task coverage comes before elegance

A beautifully written summary that omits one required task is incomplete. Copy the task verbs from the problem statement – estimate, design, compare, recommend, test – and check that each appears either directly or through an unmistakable result sentence.

3 Four functional blocks – with useful flexibility

The original draft's four-block idea is useful, but it should be treated as a design tool rather than a rigid paragraph law.

Recommended four-block architecture

Block	Main job	Must contain
1. Task	Orient the reader	Decision context, required outputs, important constraints
2. Strategy	Explain the modeling architecture	Baseline/refinement sequence, named methods, data role
3. Findings	Deliver the answer	Headline result first, units, comparison, task coverage
4. Trust and decision	Calibrate confidence and action	Validation/sensitivity evidence, threshold or caveat, recommendation

A typical one-page summary may devote roughly 2–3 sentences to Task, 3–5 to Strategy, 4–6 to

Findings, and 2–4 to Trust/Decision. These are drafting ranges, not rules.

Task: name the decision, outputs, and constraints – not the contest history.

Strategy: show how the models connect – not a list of software and equations.

Findings: lead with the answer, then the improvement, threshold, or uncertainty.

Trust/Decision: state the check, the tested range, and the action that follows.

Figure 2: A stable drafting architecture. A results-first opening is also possible when the answer is urgent and immediately interpretable.

Model criticism: When a results-first opening works

A summary may open with the answer when the problem is strongly decision-oriented:

> *We recommend four buses on ordinary school days, an early fleet review at 236 projected riders, and a fifth contingency bus from 241 confirmed riders.*

The next sentence must then identify the task and conditions. Results-first is not permission to remove context; it changes the order of evidence.

4 Writing the Task and Strategy blocks

4.1 Task block: compress the question, not the background

A strong Task block should make the required deliverable visible. It does not need a history lesson unless the background changes the modeling choice.

Worked example 2: weak and improved Task blocks

Weak

> *School buses are important in many places, and many students use them every day. This paper studies school bus routing, which is a complicated problem.*

Improved

> *The school must transport 226 students with buses of capacity 60 while reducing ride time and keeping every scheduled journey below 45 minutes. We therefore seek the smallest feasible fleet, a set of capacity-respecting routes, and a contingency rule for uncertain daily demand.*

The improved version names the decision variables, constraints, and outputs in two sentences.

Task-block sentence frame

[Decision maker] must [decision] for [system/population] under [main constraints]. We estimate/design/compare [required outputs], with [performance metric] as the primary criterion.

4.2 Strategy block: architecture before technical detail

The Strategy block should explain why multiple methods exist and how they hand off to one another. A software list is not a modeling strategy.

Worked example 3: method list versus model architecture

Weak

> *We used Python, clustering, Dijkstra's algorithm, optimization, simulation, and sensitivity analysis.*

Improved

> *We first derive a fleet-size lower bound from capacity, then construct feasible routes by clustering nearby stops and solving shortest paths on a time-weighted road graph. A local-search stage exchanges stops between routes to reduce the longest ride rather than total distance alone. Finally, we rerun the assignment under perturbed district demand to identify when four buses cease to be sufficient.*

The improved version presents a sequence of modeling decisions. Software is implementation detail and may be omitted from the Summary.

Checkpoint: Checkpoint: name the repair

For each transition, state what weakness is repaired:

1. capacity lower bound → feasible routing;
2. distance-minimizing routing → longest-ride local search;
3. nominal demand → perturbed-demand reruns.

A refinement should exist because the baseline misses something, not because a second algorithm sounds advanced.

5 Writing findings that carry evidence

The Findings block is not a storage place for every output. It should contain a hierarchy of claims.

Three levels of result

1. **Decision result:** the route, number, ranking, threshold, or recommendation the problem requested.
2. **Performance evidence:** improvement against a baseline or requirement.
3. **Robustness boundary:** the tested range within which the decision holds, and the point at which it changes.

A summary usually needs all three levels, but only a few carefully chosen numbers.

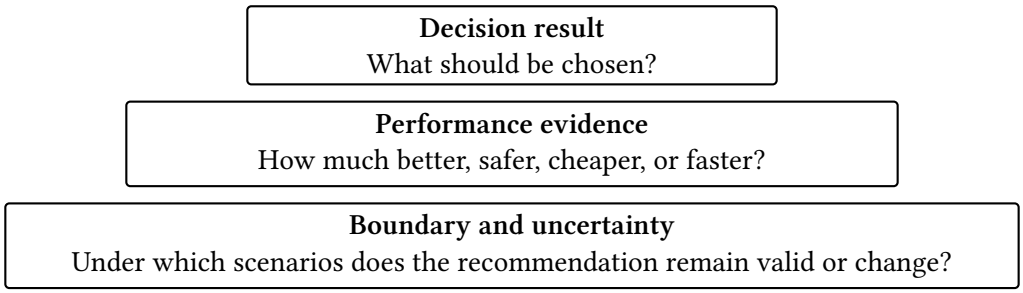

Figure 3: Result selection should move from the decision to its quantitative support and then to its tested boundary.

A quantitative claim needs context

A useful result sentence normally includes several of the following:

- **value and unit**: 28.2 minutes, 87.4 km, four buses;
- **baseline or target**: down from 42.9 minutes; below the 45-minute limit;
- **scope**: morning route, 226 registered students, ordinary-day demand;
- **uncertainty or tested range**: 18 of 20 scenarios; ± 10 students by district;
- **decision implication**: keep four buses; activate a fifth under the trigger condition.

Worked example 4: progressively stronger result sentences

Version	Sentence
Too vague	Our optimized routes are much better than the current routes.
Value only	The mean ride time is 28.2 minutes.
Compared	The mean ride time falls from 42.9 to 28.2 minutes, a 34.3% reduction.
Decision-ready	With four buses, the mean ride time falls from 42.9 to 28.2 minutes and the longest route is 44.1 minutes, meeting the 45-minute requirement.
Calibrated	With four buses, the mean ride time falls from 42.9 to 28.2 minutes and the longest route is 44.1 minutes; four buses remain feasible in 18 of 20 perturbed-demand scenarios, while two high-demand scenarios require a fifth bus.

Model criticism: Number economy

More numbers do not automatically create more evidence. Select approximately three to six headline numbers that serve different jobs: the answer, the improvement, the constraint, and the robustness boundary. Move route-by-route details, parameter tables, and secondary metrics to the body.

Checkpoint: Checkpoint: repair the denominator

"The method succeeds 90% of the time" is incomplete. Rewrite it by specifying: succeeds at what, over how many scenarios or replications, under what input range, and relative to which threshold.

6 Trust language without overclaiming

The final block should tell the reader why the findings deserve confidence and where that confidence ends.

Four distinct evidence verbs

Term	Question answered	Summary language
verification	Did we implement or solve the model correctly?	"A hand-calculated four-stop case matched the program output."
validation	Does the model reproduce relevant reality or an accepted benchmark?	"Predicted ride times differed from two held-out mornings by a mean of 2.4 minutes."
sensitivity	How do results change when inputs or assumptions change?	"The four-bus plan remained feasible in 18 of 20 perturbed-demand scenarios."
uncertainty	What range or probability surrounds the estimate?	"The simulated 90th-percentile wait was 7.8–9.1 minutes across replications."

Do not use "validated" when the only check was that code ran without errors. Do not use "robust" without naming the tested range and outcome metric.

Worked example 5: caveat without self-sabotage

Weak

> *Our model has many unrealistic assumptions, so the results may not be accurate.*

Improved

> *The four-bus recommendation is stable under the tested ± 10-student district perturbations, but it excludes road closures and unusually concentrated absences. We therefore recommend four scheduled buses, an early fleet review at 236 projected riders, and a fifth-bus contingency from 241 confirmed riders or when a major route becomes unavailable.*

The improved version identifies the tested domain, the untested condition, and the operational response.

Precision language bank

Avoid	Use when supported
Our model proves...	Our model predicts... / Under the stated assumptions, the analysis implies...
The solution is optimal.	The solution is optimal for the stated linear model. / It was the best found among the tested candidates.
The result is robust.	The recommendation did not change when p varied from 0.8 to 1.2 times its baseline value.
The model is accurate.	The mean absolute validation error was 2.4 minutes on the held-out observations.
The method is efficient.	Runtime fell from 48 to 7 seconds for the 250-node test case.

7 Complete worked transformation

This section turns the evidence inventory from Worked Example 1 into a complete competition-style Summary. The case is illustrative; the purpose is sentence design, not the bus-routing mathematics.

Poor first draft

This paper studies a school bus routing problem. We use clustering, shortest paths, an optimization algorithm, Python, and sensitivity analysis. We find that our model works well and can reduce the routes. The results show that four buses are good, and the school should use our model. There are some limitations, but future work could improve them.

Model criticism: Diagnosis

The draft mentions a topic but not the actual constraints. It lists tools but not the modeling architecture. It gives no units, baseline, threshold, or task-by-task result. "Works well," "good," and "limitations" are unsupported labels. The recommendation has no trigger condition.

Final worked Summary

TASK. The school must transport 226 registered students using buses of capacity 60 while reducing travel time and keeping every scheduled morning ride below 45 minutes. We determine the minimum feasible fleet, construct a route plan, and identify when demand uncertainty requires additional capacity.
STRATEGY. We first obtain a capacity lower bound, showing that at least four buses are necessary. We then cluster nearby stops and connect them through time-weighted shortest paths on the road network. Because minimizing total distance can leave one excessively long route, a local-search stage exchanges stops between buses to reduce the maximum ride while preserving capacity. Finally, we rerun the assignment under twenty perturbed-demand scenarios in which each district gains or loses up to ten riders.
FINDINGS. We recommend four ordinary-day routes totalling 87.4 km. Relative to the existing 103.6-km plan, total route length falls by 15.6%. The predicted mean student ride falls from 42.9 to 28.2 minutes, while the longest scheduled ride is 44.1 minutes, below the 45-minute requirement. Four buses remain feasible in 18 of 20 perturbed-demand scenarios; in the two most concentrated high-demand cases, at least one route exceeds capacity and a fifth bus is required.
TRUST AND DECISION. A hand-traced small instance reproduces the route-cost calculation, and modeled ride times are compared with recorded morning travel times before optimization. The plan is therefore suitable for ordinary demand, but it should not be treated as unconditional. We recommend scheduling four buses, beginning an early fleet review at 236 projected riders, and activating a fifth-bus contingency from 241 confirmed riders, when district demand becomes unusually concentrated, or when a major road closure invalidates the travel-time matrix.

Checkpoint: Sentence-function audit

For the final Summary, underline:

1. the sentence that states the headline answer;
2. the sentence that establishes improvement against a baseline;
3. the sentence that reports the constraint test;
4. the sentence that reports sensitivity;
5. the sentence that changes the recommendation from unconditional to conditional.

If two different required tasks rely on the same vague sentence, add specificity.

8 Compression and revision

A one-page limit creates a selection problem. The goal is not to shorten every sentence equally; it is to remove low-value material while protecting the evidence chain.

Three-pass revision

1. **Completeness pass:** Does every major task have an answer? Are the recommendation and important conditions explicit?
2. **Evidence pass:** Does each major claim have a number, comparison, check, or tested range?
3. **Compression pass:** Remove history, process narrative, duplicated method detail, generic adjectives, software names, equations, and future-work filler.

Do not begin with compression. A short incomplete summary is still incomplete.

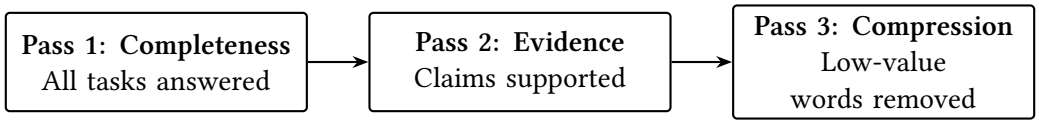

Figure 4: The revision order matters. Cutting before checking task coverage can delete the only sentence that answers a requirement.

Worked example 6: compression without losing evidence

Before – 44 words

> *In order to determine whether our proposed routing model would continue to function effectively when there were changes in the number of students assigned to each of the different geographic districts, we carried out a sensitivity analysis involving a number of alternative demand situations.*

After – 19 words

> *We reran the assignment under twenty scenarios with district demand perturbed by up to* ± 10 *students.*

The shorter sentence preserves the action, sample size, changed input, and perturbation range.

Model criticism: High-value details to protect

Protect task answers, decision thresholds, units, baselines, validation metrics, tested ranges, and recommendation conditions. Cut contest-year references, team workflow, code language, equations, exhaustive parameter lists, and generic phrases such as "future work could improve the model."

9 Writing studio and peer audit

Core route: evidence-to-summary studio

Using the school-bus evidence inventory, write a 12–16 sentence summary without copying the worked version. Requirements:

- answer all four tasks;
- include at least four units or quantitative comparisons;
- include one verification or validation sentence;
- include one sensitivity boundary;
- end with a conditional recommendation.

Challenge route: two openings

Write two versions of the first four sentences:

1. a task-first opening;
2. a results-first opening.

Exchange drafts with a partner. Which version makes the answer easier to identify without making the context ambiguous?

Extension route: summary under conflicting metrics

Suppose the shortest route plan reduces total distance to 82 km but gives one student a 51-minute ride, while the balanced plan uses 87.4 km and keeps all rides below 45 minutes. Write two sentences explaining why the balanced plan is recommended. Name the objective, the constraint, and the trade-off without calling either solution "best" in an undefined sense.

Two-minute peer audit – classroom tool, not an official rubric

Score each item 0, 1, or 2.

Criterion	Question
Task coverage	Can every major requirement be matched to an answer?
Answer visibility	Can the headline recommendation be found in ten seconds?
Model architecture	Does the reader understand how the methods connect and why refinements exist?
Quantitative evidence	Do key claims include units, baselines, thresholds, or uncertainty?
Trust calibration	Are verification/validation/sensitivity stated accurately, with a boundary or caveat?
Compression and consistency	Is every sentence useful, and do all numbers agree with the body?

After scoring, give only two comments: the highest-value sentence and the single most important repair.

Checkpoint: Exit ticket

Write one sentence for each prompt:

1. The headline answer to a modeling problem you have solved.
2. The baseline or threshold that makes the answer meaningful.
3. The evidence that supports trust in the answer.
4. The condition under which the recommendation would change.

These four sentences form the core of a future Results and Trust/Decision block.

10 What to carry into W3

Class W2 take-aways

1. Build an evidence inventory before writing prose.
2. Treat the four blocks as functions: Task, Strategy, Findings, Trust/Decision.
3. Put the headline answer before secondary outputs.
4. Report numbers with units, baselines, thresholds, and scope.
5. Use verification, validation, sensitivity, and uncertainty accurately.
6. A limitation is useful when it produces a condition, trigger, or caution.
7. Revise in the order completeness → evidence → compression.
8. The Summary must agree exactly with the body, tables, and conclusion.

In **W3: Assumptions, Validation, and Sensitivity** we develop the evidence behind the trust sentences introduced here. Students will design assumptions that can be tested, distinguish model verification from real-world validation, and turn parameter changes into decision-relevant sensitivity conclusions.

Appendix A: Evidence-inventory worksheet

Task	Headline answer	Method	Evidence/check	Condition/caveat

Appendix B: Sentence frames

- **Task:** *[Decision maker] must [decision] under [constraints]; we therefore estimate/design/compare [outputs].*

- **Architecture:** *We first [baseline], then [refinement] to address [named weakness], and finally [test/decision stage].*
- **Headline result:** *We recommend [action/quantity], which produces [metric with units] under [scope].*
- **Comparison:** *Relative to [baseline], [metric] changes from [old] to [new], a [percentage or absolute] improvement.*
- **Validation:** *Against [held-out data/benchmark], the model achieved [error or agreement metric].*
- **Sensitivity:** *The recommendation remained unchanged when [input] varied over [range], but changed when [threshold].*
- **Conditional decision:** *Use [action A] under [condition]; switch to [action B] when [trigger].*

Appendix C: Final consistency checklist

- Every required task appears in the Summary and body.
- The headline numbers match the final results table exactly.
- Units, denominators, time horizons, and scenario counts are visible.
- "Optimal," "validated," "robust," and "accurate" are supported by the correct evidence.
- No software name appears unless it matters to reproducibility or interpretation.
- No equation is included unless the equation itself is a headline result.
- The recommendation includes a decision maker, action, and relevant condition.
- The Summary contains no citation or claim absent from the reference list or body.
- The page follows the current official contest template and identification rules.

Official-source note. Contest-specific statements were checked in July 2026 against the COMAP 2026 HiMCM/MidMCM Instructions and the 2026 IM^{2}C USA Regional Instructions. Rules, templates, and regional requirements may change; verify the current official pages before each contest.

Model audit card

Audit before trusting or communicating a model

Purpose	What question or decision does the model support?	**Data**	Which inputs are observed, derived, or simulated?
Assumptions	Which simplifications may change the answer?	**Units**	Are dimensions and scales consistent?
Verification	Were equations, code, constraints, and limiting cases checked?	**Validation**	Is performance tested against independent evidence?
Uncertainty	How do measurement, parameter, solver, and structural errors affect conclusions?	**Ethics**	Who benefits, who bears risk, and what fairness criterion matters?
Reproducibility	Are data provenance, versions, seeds, and transformations recorded?	**Communication**	Are claims conditional, traceable, and decision-relevant?

Further reading

- HiMCM / MidMCM Contest Instructions (COMAP)
- Structure Your Article (IEEE Author Center)

Sources, provenance, and licence

These notes follow the iterative modeling and assessment perspective in the SIAM/COMAP GAIMME report. Competition-specific remarks are a snapshot checked on 20 August 2026 against the official COMAP HiMCM instructions; readers should verify current rules before submission. Unless a source is explicitly identified, classroom datasets and numerical scenarios are synthetic or simulated teaching examples and are not empirical claims about a real institution. Third-party references retain their own terms.

Student response guide and checking criteria

For closed calculations, show the definition, substitution, units, and at least one independent check; the worked examples in this unit are method hints, but you must complete the arithmetic yourself. Open modeling responses are checked on five dimensions: a clear purpose and output; stated data and provenance; defensible assumptions and scope; traceable computational verification and model validation; and alignment among uncertainty, limitations, and the recommendation. Support each dimension with concrete evidence rather than writing only “completed.”