Make thinking visible.

Proposed formative observation framework for ages 9–11; adaptable for younger learners. The seven indicators are not a validated psychometric scale.

Code / competency	Observable behaviour
IT / Independent thinking	Attempts an explanation before seeking help.
Q / Questioning	Asks a relevant question about a claim or its basis.
SS / Source selection	Chooses a suitable source and explains why.
ES / Evidence seeking	Identifies or checks independent evidence.
R / Reasoning	Connects evidence to a judgement.
BR / Belief revision	Maintains or revises a judgement for a reason.
IJ / Independent judgement	Explains a conclusion or defers when evidence is insufficient.

Scoring anchors

0 Not demonstrated despite opportunity. **1** Substantial support. **2** Brief prompt. **3** Independently. **N/O** No opportunity; never score as zero.

Record the task, support and evidence reference. Accept explanations through speech, drawing or examples. A correct vote alone is not sufficient; a justified uncertain judgement can be appropriate.

Stop, Ask, Check and Decide guide pupils as flexible prompts. The seven thinking competencies help teachers observe and nurture independent, evidence-informed thinking in any lesson. They are not seven additional steps.

Plan comparable observations.

A proposed evaluation protocol, not a statement that these procedures have already occurred.

Design and timing

Use comparable, age-appropriate tasks at PRE (before the teaching sequence), POST (after the sequence), and TRANSFER (a new task without AI). Record actual dates, task versions, opportunities and implementation duration. Define the timing before collection; avoid repeating identical answers.

Scoring and quality assurance

Train observers with shared examples and independent practice ratings. Specify rules for prompts and missing observations, double-score a subset and review agreement (e.g., weighted kappa where appropriate). Keep original ratings and reconciliation records. Do not combine the seven indicators into a total without measurement justification.

Analysis plan

Report valid n, missing n, median and quartiles per competency and time; means and SD are supplementary. Use Friedman on complete triplets, Kendall W for effect size, and two-sided paired Wilcoxon on complete pairs. Apply Holm correction separately to the seven omnibus and 21 pairwise tests; report ties, missingness, effect sizes and exact denominators. No imputation in the current workflow.

Interpretation

Distinguish implementation, changes in ratings and causal effects. Without a documented comparison design and validated measures, avoid causal conclusions. An omnibus difference does not identify which pair differs. Report uncertainty, limitations and null findings.

Protect the learner.

Recommended procedures for the school and research team; this document does not assert ethics approval, consent or legal compliance.

Before collecting data

Obtain appropriate school permission and independent ethics review where required. Explain participation to guardians and children in accessible language; obtain appropriate consent and assent. Participation and withdrawal should not affect learning opportunities. Avoid unnecessary personal details.

During learning

Use teacher-mediated, age-appropriate sources. No student AI accounts are required. Never send identifiable child data to AI services. Avoid public ranking, shaming, or equating language proficiency with reasoning. Offer accessible ways to demonstrate thinking.

Data stewardship

Use coded learner IDs, keep the identity key separately under restricted access, and document retention and deletion. Store original observation records securely. Record the assessor, task version and date privately. Use aggregate reporting; suppress small subgroups and avoid releasing pupil-level files.

Before public release

Verify every claimed observation against original records, review missingness, check permissions and review interpretation. Preserve source provenance. Distinguish technical test outputs from verified classroom evidence. Obtain school approval for the final aggregate report.

Release checklist: Original evidence reconciled; assessment procedures documented; consent/assent and review requirements met; aggregate outputs checked; school approval recorded.