

# MMR1: Enhancing Multimodal Reasoning with Variance-Aware Sampling

## Supplementary Material

### A. Variance–Progress Theory

**Intuition.** Gradient updates are informative only if sampled rollouts produce *different* rewards. When rewards are nearly identical, advantages collapse and gradients vanish. Our theorem shows that (under standard smoothness and non-degeneracy conditions) the expected improvement after a single update is *linearly* lower-bounded by the reward variance of the prompt. Thus, selecting prompts that induce mixed outcomes (OVS) and diverse reasoning paths (TDS) provably strengthens learning.

At a glance (notation & assumptions). We write  $g(x, y) = \nabla_{\theta} \log \pi_{\theta}(y | x)$ ,  $\bar{R}(x) = \mathbb{E}_y[R(x, y)]$ ,  $R_{\text{res}} = R - \bar{R}$ , and  $\Gamma_{\theta}(x) = \mathbb{E}_y[g g^{\top}]$ . Assumptions: (i)  $L$ -smoothness of  $J_x$ ; (ii) bounded score function  $\|g\| \leq G_{\max}$ ; (iii) uniform positive definiteness of  $\Gamma_{\theta}(x) \succeq \lambda_{\min} I$ ; (iv) gradient lower bound  $\|\nabla J_x\|^2 \geq c_{\min} \text{Var}[R]$  (a mild consequence of (iii)).

#### A.1. Technical Preliminaries

Throughout we fix a prompt  $x \in \mathcal{X}$  and write expectations over  $y$  as  $\mathbb{E}_y[\cdot] = \mathbb{E}_{y \sim \pi_{\theta}(\cdot | x)}[\cdot]$ . The policy is differentiable and strictly positive on  $\mathcal{Y}$ . We denote  $\bar{R}(x) = \mathbb{E}_y[R(x, y)]$  and  $R_{\text{res}}(x, y) = R(x, y) - \bar{R}(x)$ .

##### A.1.1. Action-independent baselines

**Proposition 1** (Optimal action-independent baseline). *For any square-integrable reward  $R$ , the baseline  $b^*(x) = \bar{R}(x)$  minimizes the total variance of the REINFORCE gradient estimator over all baselines  $b(x)$  that depend on  $x$  but not on  $y$ .*

*Proof.* Let  $G_b(x) = \mathbb{E}_y[g(x, y)(R(x, y) - b(x))]$ , where  $g(x, y) = \nabla_{\theta} \log \pi_{\theta}(y | x)$ . Using  $\mathbb{E}_y[g] = 0$ , the covariance term vanishes and  $\text{Var}[G_b(x)] = \mathbb{E}_y[\|g\|^2 (R - b)^2]$ . The right-hand side is a convex quadratic in  $b$ . Differentiating and setting to zero yields  $b(x) = \bar{R}(x)$  [77].  $\square$

**Remark.** An action-dependent baseline such as  $b^{\|g\|^2}(x) = \mathbb{E}_y[\|g\|^2 R] / \mathbb{E}_y[\|g\|^2]$  can further reduce scalar variance [12], but requires inner Monte-Carlo estimates. The bounds below hold for any  $y$ -independent baseline.

##### A.1.2. Gradient-variance bounds

With  $b^*(x) = \bar{R}(x)$ ,

$$G(x) = \mathbb{E}_y[g(x, y)R_{\text{res}}(x, y)], \mathbb{E}[G(x)] = \nabla_{\theta} J_x(\theta).$$

Its covariance is

$$\text{Var}[G(x)] = \mathbb{E}_y[R_{\text{res}}^2 g g^{\top}] - \nabla_{\theta} J_x(\theta) \nabla_{\theta} J_x(\theta)^{\top}.$$

For bounding we drop the nonnegative outer product, which only reduces the variance, yielding a valid but looser inequality.

**Assumption 2** (Bounded score-function gradient). *There exists  $G_{\max} < \infty$  such that  $\|g(x, y)\| \leq G_{\max}$  for all  $(x, y)$ .*

**Lemma 1** (Variance sandwich bound). *Under Assumption 2, let  $\Gamma_{\theta}(x) = \mathbb{E}_y[g(x, y)g(x, y)^{\top}]$ . If  $\lambda_{\min} > 0$  is the smallest eigenvalue of  $\Gamma_{\theta}(x)$  uniformly over  $x$ , then*

$$\begin{aligned} \lambda_{\min} \text{Var}_y[R(x, y)] I_d &\preceq \text{Var}[G(x)] \\ &\preceq G_{\max}^2 \text{Var}_y[R(x, y)] I_d. \end{aligned}$$

Thus reward variance is the only prompt-dependent factor; Fisher terms contribute bounded, model-dependent constants.

*Proof.* For any unit vector  $v$ ,  $v^{\top} \text{Var}[G(x)]v = \mathbb{E}_y[(v^{\top} g)^2 R_{\text{res}}^2]$ . Upper bound:  $(v^{\top} g)^2 \leq \|g\|^2 \leq G_{\max}^2$ . Lower bound:  $\Gamma_{\theta}(x) \succeq \lambda_{\min} I$  implies  $v^{\top} \text{Var}[G(x)]v \geq \lambda_{\min} \mathbb{E}_y[R_{\text{res}}^2] = \lambda_{\min} \text{Var}_y[R]$ .  $\square$

##### A.1.3. Bounds on the Fisher term

**Proposition 2** (Uniform Fisher bounds). *Under Assumption 2, there exist constants  $0 < \lambda_{\min} \leq \lambda_{\max} = G_{\max}^2$  such that*

$$\lambda_{\min} I_d \preceq \Gamma_{\theta}(x) \preceq \lambda_{\max} I_d, \quad \forall x \in \mathcal{X}.$$

*Proof.* The upper bound follows from  $\|g\| \leq G_{\max}$ . The lower bound holds under the standard non-degeneracy assumption that  $\pi_{\theta}(\cdot | x)$  defines a full-dimensional exponential family and  $\theta$  ranges over a compact set, ensuring  $\Gamma_{\theta}(x) \succ 0$  uniformly [2].  $\square$

### A.2. Proof of the Variance–Progress Theorem

Let the update be  $\theta^+ = \theta + \eta G(x)$  with  $\eta > 0$ . A second-order Taylor expansion yields

$$J_x(\theta^+) - J_x(\theta) = \eta \langle \nabla J_x, G(x) \rangle + \frac{1}{2} \eta^2 G(x)^{\top} H_x(\tilde{\theta}) G(x),$$

for some  $\tilde{\theta}$  on the segment  $[\theta, \theta^+]$ . Taking expectations and  $L$ -smoothness,

$$\mathbb{E}[J_x(\theta^+) - J_x(\theta)] \geq \eta \|\nabla J_x\|^2 - \frac{1}{2} \eta^2 L \mathbb{E}[\|G(x)\|^2].$$

Decomposing  $\mathbb{E}[\|G(x)\|^2]$  into bias and variance and using Lemma 1,

$$\begin{aligned} \mathbb{E}[J_x(\theta^+) - J_x(\theta)] &\geq \eta \|\nabla J_x\|^2 \\ &\quad - \frac{1}{2} \eta^2 L (\|\nabla J_x\|^2 + d G_{\max}^2 \text{Var}[R]). \end{aligned}$$

Assuming  $\|\nabla J_x\|^2 \geq c_{\min} \text{Var}[R]$ ,

$$\begin{aligned} \mathbb{E}[J_x(\theta^+) - J_x(\theta)] &\geq \eta c_{\min} \text{Var}[R] \\ &\quad - \frac{1}{2} \eta^2 L (c_{\min} + d G_{\max}^2) \text{Var}[R]. \end{aligned}$$

Choosing  $\eta \leq \frac{c_{\min}}{2L(c_{\min} + dG_{\max}^2)}$  gives  $\mathbb{E}[J_x(\theta^+) - J_x(\theta)] \geq \frac{\eta c_{\min}}{2} \text{Var}[R]$ . Restricting further to  $\eta \leq c_{\min}/(4L)$  yields the simplified bound used in the main text:

$$\mathbb{E}[J_x(\theta^+) - J_x(\theta)] \geq \frac{\eta c_{\min}}{4} \text{Var}_{y \sim \pi_\theta}[R(x, y)].$$

**Discussion.** The bound depends only on constants  $c_{\min}$  and  $L$  tied to the model family. All prompt dependence enters through  $\text{Var}[R]$ , establishing reward variance as the decisive quantity.

### A.3. Two-Level Decomposition of Reward Variance

Let  $y = (y_{\text{cot}}, y_{\text{ans}})$  and define  $R(x, y) = \mathbf{1}_{\text{verifier}(y_{\text{ans}})}$ . Let  $Z = \varphi(y_{\text{cot}})$  represent the chain. By the law of total variance,

$$\text{Var}_y[R] = \mathbb{E}_Z[p_Z(1 - p_Z)] + \text{Var}_Z[p_Z],$$

where  $p_Z(x) = \Pr(R = 1 \mid Z)$ .

**Intra/inter-trajectory terms.** The first term is intra-trajectory Bernoulli variance; the second is inter-trajectory variation of success probabilities.

**Efron–Stein lower bound.** If  $|p_z - p_{z'}| \leq L d(y_{\text{cot}}(z), y_{\text{cot}}(z'))$  for a bounded distance  $d \in [0, 1]$ , then

$$\text{Var}_Z[p_Z] \geq \frac{L^2}{4} \mathbb{E}_{Z, Z'}[d^2(y_{\text{cot}}(Z), y_{\text{cot}}(Z'))].$$

**OVS and TDS estimators.** With  $K$  rollouts,  $\hat{p} = \frac{1}{K} \sum R$ ,

$$\begin{aligned} \widehat{\text{OVS}}(x) &= \hat{p}(1 - \hat{p}), \\ \widehat{\text{TDS}}(x) &= \frac{1}{K(K-1)} \sum_{i \neq j} d^2(y_{\text{cot}}^{(i)}, y_{\text{cot}}^{(j)}). \end{aligned}$$

By the strong law and U-statistic convergence, both estimators are strongly consistent for their respective population terms.

**Variance Promotion Score (VPS).** Define  $\widehat{\text{VPS}} = \alpha \widehat{\text{OVS}} + \beta \widehat{\text{TDS}}$  with  $\alpha, \beta > 0$ . Then  $\widehat{\text{VPS}}$  converges almost surely to a positive affine transform of a *lower bound* on  $\text{Var}[R]$ . Hence, VPS is a strongly consistent monotone surrogate for reward variance (it does not require equality to hold).

### A.4. Extension from REINFORCE to GRPO

GRPO replaces the scalar baseline by the in-batch mean reward and whitens with the sample standard deviation:

$$\tilde{R}(x, y_i) = \frac{R(x, y_i) - \bar{r}(x)}{s(x) + \delta}.$$

This yields a centered, variance-controlled REINFORCE estimator. Multiplying by importance ratios (and optionally clipping) preserves the core dependence on reward variance: the Variance–Progress lower bound continues to hold after a rescaling of constants. When clipping is enabled, an  $O(\varepsilon)$  bias arises; the bound is reduced by the same order but remains strictly positive whenever  $\text{Var}[R] > 0$ . Thus prompts that induce higher reward variance guarantee larger expected improvements under GRPO, paralleling vanilla REINFORCE.

## B. Fine-Grained Math Type Definitions

For the construction of our math dataset, each problem is assigned to one of thirteen fine-grained categories. These categories provide balanced coverage across fundamental and advanced domains of mathematics, and ensure that the RL dataset spans diverse reasoning skills. The formal definitions, explanations, and illustrative examples for each category are presented below.

### 1. Arithmetic

*Definition:* Arithmetic covers basic numerical operations, including addition, subtraction, multiplication, and division.

*Explanation:* It forms the foundation of mathematics by establishing rules for manipulating numbers.

*Example:* Compute  $15 \times 12$ .

### 2. Counting

*Definition:* Counting addresses enumeration of objects or elements within a collection.

*Explanation:* It includes both simple enumeration and principles such as permutations and combinations.

*Example:* Determine how many integers between 1 and 100 are multiples of 5.

### 3. Combinatorics

*Definition:* Combinatorics studies arrangements, selections, and combinations of discrete objects.

*Explanation:* It extends counting principles to complex scenarios involving structured sets.

*Example:* How many ways can 5 distinct books be arranged in a row?

### 4. Algebra

*Definition:* Algebra represents relationships using symbols and equations.

*Explanation:* It provides systematic tools for solving equations and reasoning about unknowns.

*Example:* Solve  $2x + 3 = 9$  for  $x$ .

### 5. Functions

*Definition:* A function maps each input to exactly one output.

*Explanation:* Functions formalize dependencies between quantities, described via formulas, graphs, or rules.

*Example:* Given  $f(x) = x^2 + 3$ , find  $f(2)$ .

### 6. Plane Geometry

*Definition:* Plane Geometry studies figures such as lines, angles, and polygons in two dimensions.

*Explanation:* It addresses lengths, angles, and areas of flat figures.

*Example:* Find the area of a triangle with base 10 and height 5.

#### 7. Solid Geometry

*Definition:* Solid Geometry extends geometric reasoning to three-dimensional figures.

*Explanation:* It concerns volumes, surface areas, and properties of 3D objects.

*Example:* Find the volume of a cube with side length 4.

#### 8. Combinatorial Geometry

*Definition:* Combinatorial Geometry analyzes discrete configurations of geometric objects.

*Explanation:* It merges counting with geometry, such as enumerating diagonals or intersections.

*Example:* How many diagonals does a convex octagon have?

#### 9. Descriptive Geometry

*Definition:* Descriptive Geometry represents 3D objects using 2D projections.

*Explanation:* It enables precise measurement of spatial relationships via orthographic or perspective drawings.

*Example:* Sketch the top and front views of a cube resting on a horizontal plane.

#### 10. Graph Theory

*Definition:* Graph Theory studies structures composed of vertices and edges.

*Explanation:* It focuses on connectivity, paths, and cycles in discrete networks.

*Example:* Given a graph with vertices A, B, C, D and edges AB, BC, CD, and DA, does the graph contain a cycle?

#### 11. Logic

*Definition:* Logic studies formal reasoning and inference.

*Explanation:* It examines propositions, truth values, and the validity of conclusions.

*Example:* If "All cats are mammals" and "Fluffy is a cat," deduce whether "Fluffy is a mammal."

#### 12. Statistics

*Definition:* Statistics concerns the collection, analysis, and interpretation of data.

*Explanation:* It uses measures such as mean, median, and variance to summarize data.

*Example:* For  $\{2, 4, 4, 6, 8\}$ , compute the mean and median.

#### 13. Topology

*Definition:* Topology studies properties of spaces invariant under continuous deformations.

*Explanation:* It focuses on qualitative properties such as connectivity and the number of holes.

*Example:* Explain why a doughnut (torus) cannot be de-

formed into a sphere without removing the hole.

These categories are used to uniformly sample math problems in the RL stage, ensuring both difficulty balance and coverage across diverse mathematical skills.

## C. Detailed Hyper-parameters

The main hyperparameters used in the cold-start supervised fine-tuning stage are summarized in Table 5.

Setting	Value
Training epochs	3
Gradient accumulation steps	4
Effective batch size	32
Sequence length cutoff	16,384
Optimizer	AdamW
Learning rate	$1 \times 10^{-5}$
Learning rate schedule	Cosine decay
Warm-up ratio	0.1
Weight decay	0
Max gradient norm	1.0
Precision	bfloat16
Deepspeed config	ZeRO-3 (32 shards)

Table 5. Cold-start stage hyperparameters.

The main hyperparameters used in the RL stage (GRPO with VAS) are summarized in Table 6.

## D. System Prompt

**Training Prompt.** The following system prompt is used during RL training to enforce a structured reasoning format and to require the final answer to be enclosed in “\boxed{ }”.

**Evaluation Prompt.** To ensure fairness and generalizability, we use a simplified prompt for evaluation instead of the training prompt.

## E. Qualitative Demonstration

As shown in Figure 4, this MathVerse question highlights the reasoning capability of MMR1. The solution it generates follows a clear and logical structure: the model begins by restating the given conditions, then applies the angle-sum property of a triangle to determine the missing angle, and finally uses parallel-line properties to compute the target angle. This step-by-step organization reflects a coherent “analyze–plan–execute” process. The response also demonstrates reflective reasoning. After deriving the result, the model verifies consistency with geometric rules and cross-checks against the answer choices. Furthermore, it provides

Setting	Value
Initialization (policy)	Cold-start checkpoint (Qwen2.5-VL)
Objective	GRPO (adv_estimator=grpo)
Reward	1 * format reward + 1 * accuracy reward
KL regularization	Enabled, penalty=low_var_kl, coef = 0.01 (fixed)
Precision	bfloat16
System prompt	As in Appx. D
<i>Data &amp; VAS sampling</i>	
Max prompt / response length	2048 / 4096
Image resolution limits	[7,056, 1,048,576] pixels
Sampling strategy	VAS (curriculum)
VAS metrics / weights	learnability (OVS) 0.8, self_bleu_123 (TDS) 0.2
VAS update frequency	56 steps
VAS mixture ratio	0.5 (weighted vs. uniform)
VPS rollout for scoring	$n = 16$ , batch size = 4096
<i>Rollout &amp; inference</i>	
Sampler	vLLM
Samples per prompt	$n = 8$
Temperature / top- $p$ / top- $k$	0.6 / 1.0 / -1
Validation override	temp = 0.5, $n = 1$
GPU memory util (vLLM)	0.75
<i>Optimization (actor / critic)</i>	
Global batch size (actor)	512
Micro-batch (update) (actor)	8
Micro-batch (experience) (actor)	32
PPO epochs	20
Clipping	policy clip_low=0.2, clip_high=0.2
Max grad norm	1.0 (both)
Optimizer	AdamW (lr = $1 \times 10^{-6}$ , betas = (0.9, 0.999))
Weight decay	0.01
LR warmup ratio	0.0 (constant warmup style)
<i>Parallelism &amp; hardware</i>	
FSDP	Full shard (policy), fsdp_size=8
Episodes	total_episodes = 10

Table 6. RL stage hyperparameters.

Table 7. System prompt used in RL training.

A conversation between User and Assistant. The User provides an image and asks a question. The Assistant first analyzes both the image and the question, then carefully thinks about the reasoning process step by step, and finally provides the User with an accurate answer. The Assistant must carefully check out the correctness and validity of each reasoning step. If any errors or inconsistencies are found during the reasoning process, the Assistant reflects and corrects them logically. The reasoning process and answer are enclosed within <think> </think> and <answer> </answer> tags, respectively, i.e., <think> reasoning process here, with potential reflections and corrections </think><answer> final answer here, with the key result enclosed in \boxed{ } </answer>.

an alternative method based on corresponding and supplementary angles, which strengthens confidence in the correctness of the conclusion. Overall, the model not only produces

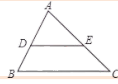
the correct answer but also exhibits robust reasoning behaviors, including systematic decomposition, verification, and multiple-solution perspectives, which illustrate strong

Table 8. System prompt used for evaluation.

Please solve the problem step by step and put your answer in one `\boxed{}`. If it is a multiple-choice question, only one letter should appear inside the `\boxed{}`.

### Question

As shown in the figure, in triangle ABC, it is known that  $\angle A = 80.0$ ,  $\angle B = 60.0$ , point D is on AB and point E is on AC, DE parallel BC, then the size of angle CED is (). Choices: A:  $40^\circ$  B:  $60^\circ$  C:  $120^\circ$  D:  $140^\circ$



### Thinking Process

1. **Analyze** the given information:
  - \* We have a triangle ABC. Angle A =  $80.0$  degrees. Angle B =  $60.0$  degrees. Point D is on side AB. Point E is on side AC. Line segment DE is parallel to line segment BC (DE  $\parallel$  BC).
2. **Identify** the goal:
  - \* Find the measure of angle CED.
3. **Plan** the solution:
  - \* Use the property that the sum of angles in a triangle is  $180$  degrees to find angle C in triangle ABC.
  - \* Use the property of parallel lines (DE  $\parallel$  BC) and transversals to relate angle CED to other angles in the figure.
4. **Execute** the plan:
  - Step 1: Find angle C in triangle ABC.**
    - \* In triangle ABC, the sum of angles is  $180^\circ$ .
    - \* Angle A + Angle B + Angle C =  $180^\circ$
    - \*  $80.0^\circ + 60.0^\circ + \text{Angle C} = 180^\circ$
    - \*  $140.0^\circ + \text{Angle C} = 180^\circ$
    - \* Angle C =  $180^\circ - 140.0^\circ$
    - \* Angle C =  $40.0^\circ$
  - Step 2: Use the parallel lines property.**
    - \* We are given that DE  $\parallel$  BC.
    - \* Consider AC as a transversal line intersecting the parallel lines DE and BC.
    - \* The angles CED and ACB (which is angle C) are consecutive interior angles (also called same-side interior angles).
    - \* Consecutive interior angles formed by parallel lines and a transversal are supplementary (their sum is  $180^\circ$ ).
    - \* Therefore, Angle CED + Angle ACB =  $180^\circ$ .
    - \* Angle CED + Angle C =  $180^\circ$ .
    - \* Substitute the value of Angle C found in Step 1:
    - \* Angle CED +  $40.0^\circ = 180^\circ$ .
    - \* Angle CED =  $180^\circ - 40.0^\circ$ .
    - \* Angle CED =  $140.0^\circ$ .
  - Alternative Step 2 (Using corresponding angles):**
    - \* Since DE  $\parallel$  BC, consider AC as a transversal.
    - \* Angle AED and Angle ACB (Angle C) are corresponding angles.
    - \* Therefore, Angle AED = Angle C =  $40.0^\circ$ .

### Thinking Process (Cont'd)

- \* Points A, E, C are collinear, forming a straight line. Angles AED and CED form a linear pair.
  - \* Angles in a linear pair are supplementary.
  - \* Therefore, Angle AED + Angle CED =  $180^\circ$ .
  - \*  $40.0^\circ + \text{Angle CED} = 180^\circ$ .
  - \* Angle CED =  $180^\circ - 40.0^\circ = 140.0^\circ$ .
5. **Verify the result:**
    - \* Both methods yield Angle CED =  $140.0^\circ$ .
    - \* The steps are logically sound and use correct geometric properties.
  6. **Compare with choices:**
    - \* The calculated value is  $140.0^\circ$ .
    - \* The choices are A:  $40^\circ$ , B:  $60^\circ$ , C:  $120^\circ$ , D:  $140^\circ$ .
    - \* The result matches choice D.
  7. **Final Answer Formulation:**
    - \* The reasoning involves calculating angle C and then using the property of consecutive interior angles formed by parallel lines DE and BC with transversal AC.
    - \* Angle C =  $180^\circ - (80^\circ + 60^\circ) = 40^\circ$ .
    - \* Angle CED + Angle C =  $180^\circ$  (consecutive interior angles).
    - \* Angle CED =  $180^\circ - 40^\circ = 140^\circ$ .
    - \* The correct option is D.

### Final Answer

1. Find the measure of angle C in triangle ABC. The sum of angles in a triangle is  $180^\circ$ .  
 $\text{Angle C} = 180^\circ - (\text{Angle A} + \text{Angle B}) = 180^\circ - (80.0^\circ + 60.0^\circ) = 180^\circ - 140.0^\circ = 40.0^\circ$
  2. Use the property of parallel lines. Since DE  $\parallel$  BC, and AC is a transversal line intersecting these parallel lines, the consecutive interior angles (also known as same-side interior angles) are supplementary.  
 $\text{Angle CED} + \text{Angle ACB} = 180^\circ$   
 $\text{Angle CED} + \text{Angle C} = 180^\circ$   
 $\text{Angle CED} + 40.0^\circ = 180^\circ$   
 $\text{Angle CED} = 180^\circ - 40.0^\circ = 140.0^\circ$
- Alternatively:
1. Find the measure of angle C in triangle ABC: Angle C =  $40.0^\circ$ .
  2. Since DE  $\parallel$  BC, and AC is a transversal, the corresponding angles are equal. Angle AED = Angle ACB = Angle C =  $40.0^\circ$ .
  3. Angles AED and CED form a linear pair on the straight line AC. Therefore, they are supplementary.  
 $\text{Angle AED} + \text{Angle CED} = 180^\circ$   
 $40.0^\circ + \text{Angle CED} = 180^\circ$   
 $\text{Angle CED} = 180^\circ - 40.0^\circ = 140.0^\circ$
- The size of angle CED is  $140.0^\circ$ .  
 Comparing this with the choices: A:  $40^\circ$ , B:  $60^\circ$ , C:  $120^\circ$ , D:  $140^\circ$ .  
 The correct option is 'boxed{D}'.

Figure 4. Qualitative demonstration of MMR1's reasoning process on a MathVerse problem. The figure illustrates the input question, the model's step-by-step thinking process, and the final answer. The reasoning is logically structured, including problem analysis, solution planning, execution, verification, and alternative approaches, ultimately arriving at the correct answer.

problem-solving ability beyond direct computation.