Table 1: Model Accuracy Comparison with Improved Formatting[a,b]

Category	Model	Setting	AIME24	GPQA	Math 500
Prompting on Instruct Models					
	Qwen 2.5-1.5B-instruct	Standard	3.33	18.97	25.00
		Zero Shot	0.00	19.20	35.80
		CoT	6.67	17.86	38.00
		CoD	0.00	20.76	26.60
	DeepSeek-R1-Distill-Qwen-1.5B	Zero Shot	3.33	5.36	45.20
Fine-tuned Models					
	Llama-3.2-1B-instruct[b]	GRPO (temp 0.0)	0.00	19.87	33.80
		GRPO (temp 0.8)	0.00	17.86	29.80
	Qwen 2.5-1.5B-instruct[a]	SFT (2e-4)	3.33	12.05	35.20
		SFT (1e-4)	0.00	13.39	29.00
		SFT (1e-5)	0.00	16.74	34.00
		SFT (5e-6)	0.00	20.54	38.00
		SFT (1e-6)	3.33	17.19	35.40
Progressive Reasoning Expansion					
	Phi-4-mini-instruct	Zero Shot	3.33	30.13	51.40
		No Wait	6.66	24.10	53.80
		1 Wait	6.66	23.88	58.00
		2 Wait	6.66	26.33	60.60
		3 Wait	10.00	27.00	64.00

[a] SFT: Fine-tuned on S1K dataset with different learning rates
[b] GRPO: Fine-tuned on GSM8K dataset with different temperature settings inferences
Note: All accuracy values are percentages. Shading indicates major category groups.

Table 2: Response Token Counts Comparison[a,b]

Category	Model	Setting	AIME24	GPQA	Math 500
Prompting on Instruct Models					
	Qwen 2.5-1.5B-instruct	Standard	762	405	852
		Zero Shot	830	807	753
		CoT	1592	775	801
		CoD	444	813	384
	DeepSeek-R1-Distill-Qwen-1.5B	R1	8308	7357	6561
Fine-tuned Models					
	Llama-3.2-1B-instruct[b]	GRPO (temp 0.0)	2948	1247	1649
		GRPO (temp 0.8)	1184	567	751
	Qwen 2.5-1.5B-instruct[a]	SFT (2e-4)	7821	6039	4418
		SFT (1e-4)	7685	4623	6130
		SFT (1e-5)	5338	2598	2357
		SFT (5e-6)	2840	1552	1544
		SFT (1e-6)	1311	796	974
Progressive Reasoning Expansion					
	Phi-4-mini-instruct	Zero Shot	687	642	462
		No Wait	784	753	567
		1 Wait	1136	916	696
		2 Wait	1418	1139	905
		3 Wait	1927	1432	1221

[a] SFT: Fine-tuned on S1K dataset with different learning rates.
[b] GRPO: Fine-tuned on GSM8K dataset with different temperature settings in inference.